

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 12:56:08 ; Search time 18 Seconds
(without alignments)
2019.536 Million cell updates/sec

Title: US-10-067-385-8

Perfect score: 4026

Sequence: 1 KLGTAESKFKVLNGKSGS.....ATVLKNNISSKTTNNPNK 773

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	219	5.4	1076	1 RPOB_ASTLO	P27059 astasia lon
2	218	5.4	1630	1 MSP1_PLAFK	P04932 plasmodium
3	218	5.4	1639	1 MSP1_PLAFW	P04933 plasmodium
4	215.5	5.4	1251	1 RBP2_PLAFV	Q00799 plasmodium
5	214	5.3	2339	1 RPL1_PLAFV	P27625 plasmodium
6	209.5	5.2	1104	1 COLA_CLOPE	P43153 clostridium
7	194	4.8	1957	1 SPOF_SCHPO	Q10411 schizosacch
8	193	4.8	2869	1 RBPI_PLAFV	Q00798 plasmodium
9	190	4.7	2452	1 RPB1_PLAFD	P14248 plasmodium
10	186.5	4.6	1790	1 USO1_YEAST	P25386 saccharomyc
11	185	4.6	3259	1 GIAN_HUMAN	Q14789 homo sapien
12	184	4.6	1875	1 MLPI_YEAST	Q02455 saccharomyc
13	183.5	4.6	1053	1 SLPM_BACBR	P06546 bacillus br
14	183	4.5	2334	1 WAPA_BACSU	Q07833 bacillus su
15	179.5	4.5	1183	1 CNA_STRAU	Q53654 staphylococ
16	178	4.4	1433	1 SUBF_BACSU	P16397 bacillus su
17	178	4.4	1726	1 MSP1_PLAFV	P04934 plasmodium
18	178	4.4	1726	1 MSP1_PLAFV	P04934 plasmodium
19	178	4.4	2492	1 ATRX_HUMAN	P50495 plasmodium
20	177	4.4	800	1 CYAA_BACAN	P46100 homo sapien
21	176.5	4.4	1116	1 SLPH_BRECH	P40136 bacillus an
22	174	4.3	735	1 CIGB_DICDI	P38538 brevivacill
23	174	4.3	1176	1 YOH8_YEAST	Q94481 dictyosteli
24	174	4.3	1462	1 GTPD_STRMU	Q08236 saccharomyc
25	174	4.3	1628	1 NAGH_CLOPE	P49331 streptococc
26	173.5	4.3	944	1 NUF1_YEAST	P26831 clostridium
27	173	4.3	2710	1 TOXA_CLODI	P32380 saccharomyc
28	172.5	4.3	596	1 PRIM_CLOAB	P16154 clostridium
29	172.5	4.3	1545	1 IGAI_HAEIN	P33655 clostridium
30	172	4.3	873	1 SYA_WIGBR	P45385 haemophilus
31	171	4.2	999	1 HGPT_HAEIN	Q8d2w8 wigleswort
32	171	4.2	1701	1 MSP1_PLAFM	P44809 haemophilus
33	170	4.2	970	1 T3RE_BPP1	P08569 plasmodium
					P08764 bacterioph

ALIGNMENTS

RESULT 1

ID	RPOB_ASTLO	STANDARD;	PRT;	1076 AA.
AC	P27059;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	DNA-directed RNA polymerase beta chain (EC 2.7.7.6).			
GN	RPOB.			
OS	Astasia longa (Euglenophyceae alga).			
OG	Chloroplast.			
OC	Eukaryota; Euglenozoa; Euglenida; Euglenales; Euglena.			
OX	NCBI_TaxID=3037;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CCAP 1204-17a;			
RX	MEDLINE=21080522; PubMed=11212895;			
RA	Gockel G., Hachtel W.;			
RT	"Complete gene map of the plastid genome of the nonphotosynthetic euglenoid flagellate Astasia longa.";			
RL	Protist 151:347-351(2000).			
RN	[2]			
RP	PRELIMINARY SEQUENCE OF 1-524 FROM N.A.			
RC	STRAIN=CCAP 1204-17a;			
RX	MEDLINE=91176556; PubMed=2078869;			
RA	Siemeister G., Buchholz C., Hachtel W.;			
RT	"Genes for ribosomal proteins are retained on the 73 kb DNA from Astasia longa that resembles Euglena chloroplast DNA.";			
RL	Curr. Genet. 18:457-464(1990).			
RN	[3]			
RP	REVISIONS.			
RC	STRAIN=CCAP 1204-17a;			
RX	MEDLINE=95163124; PubMed=7859309;			
RA	Gockel G., Hachtel W., Baier S., Fliss C., Henke M.;			
RT	"Genes for components of the chloroplast translational apparatus are conserved in the reduced 73-kb plastid DNA of the nonphotosynthetic euglenoid flagellate Astasia longa.";			
RL	Curr. Genet. 26:256-262(1994).			
CC	FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS SUBSTRATES.			
CC	CAATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +			
CC	(RNA) (N).			
CC	SUBUNIT: In chloroplasts, the RNA polymerase is composed of four subunits: alpha, beta, beta', and beta".			
CC	SUBCELLULAR LOCATION: Chloroplast.			
CC	SIMILARITY: Belongs to the RNA polymerase beta chain family.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			

34	169.5	4.2	1928	1 MYS1_YEAST	P08964 saccharomyc
35	168	4.2	1163	1 SBCC_CLOAB	Q97Fk1 clostridium
36	168	4.2	1701	1 MSP1_PLAFV	P13819 plasmodium
37	167.5	4.2	1541	1 IGAI_HAEIN	P42782 haemophilus
38	167	4.1	1451	1 SPT6_YEAST	P23615 saccharomyc
39	167	4.1	1682	1 MSP1_PLAF3	P19598 plasmodium
40	167	4.1	3305	1 APLP_VANSE	Q25490 manduca sex
41	166.5	4.1	875	1 ZIPI_YEAST	P31111 saccharomyc
42	166.5	4.1	1147	1 CGAI_HELPY	P80200 helicobacte
43	165.5	4.1	1005	1 RASO_METJA	Q58718 methanococ
44	165.5	4.1	1233	1 SMIA_MOUSE	Q9cu62 mus musculu
45	165	4.1	1176	1 SLAP_BACSH	P38537 bacillus sp

Db 223 KDVGMEDYIKKNTIENIELIEESKTKIDKKNATKBE-----KKLYQAOY 274
QY 76 FTSETGKRMEEYDYKDDRGNIAYDGDGTDEYETEKLEIKSKIYGVLSPKDGHFEIL 135
Db 275 DLSIYNKOLEE-----AHNLISV-----LE---KRIDTLK-----KNIENIKELL 310
QY 136 GKISNVSKNAKYVGNYSKIEIKATKYDFH-----SKTMTFDLYANINDIVDGLAFA 188
Db 311 DKINEIKNPPANGNTPTLLDKNKKIEHEKEIKETAKTIKFNIDSIFTDPL-----364
QY 189 GDMRLFVKNDOKKAEIKIRMEPEKIKETKSEYP-----YVSSYGNV-----IEIGE-GDLSK 239
Db 365 -ELEYLRKN-KNIDISAKVETKESTENEPNGVTYPLSYNDINNALNELNSFGDLI- 421
QY 240 NKPDNLTKMESGKIYSDSEKQOYL--LKDNIIL-----RKGYALKVTTYTPGKTMDL 289
Db 422 -NPFDTYKPSKNIYDNERKKFINEIKIEKIEKKIESDKSYEDRSKSLN---DIT 476
QY 290 EG-----NGVYSKE-----DIAKIQKA-----NPNRLALSETTIYADSRNVEDGRSTQSV 334
Db 477 KEYEKLLEIYDSKFNNDIDLTNFEKMGKRYSYKVEKLTHNTFASVENSKHNLKLTJK 536
QY 335 LMSALDGF--NII-----RQVFTFKMNDKGEAIDKGNLVTSSKLVLFGKDKKEYT 385
Db 537 AUKTYMEDYSLRNIVVEKELKYKLNLSIENETELVE--NIKOEEOQLF-----EKKIT 589
QY 386 GBDKFNVEAIKEDGSMFLDTDKPVLNSMDKNYFNPSKSNKIYVRNPEFYLRGKISDKGOF 445
Db 590 KDNKPDEKILEVSDIVKVOQVLL--MNKIDELKTKQLILKNVE--LKHNIHVPSY 644
QY 446 NMLRVNESVVDNLIYDGLHIDNTRDNFKL-----NVK-----480
Db 645 KOENKOE---PYLILVKKIEIDKLKVPKPVESLINEEKKNIKTEGQSDNSEPSTGEI 700
QY 481 -----DGDIMDMGMDYK-----ANGFPDKV 501
Db 701 TGOATTKPQOAGSALLEGDSVOAQOEOQAOQPPVPEAKAQVTPPPAPVNNKTENV 760
QY 502 TMDGNYVILQYSDINA-----KAVGHYQFLYDNVKNPEVNDPKGNTSIEYADGKSV 556
Db 761 SKLD---YLEKLYEFLNTSYCHKYLVSHTMNEKILQYKITEEESKLSLSCDPLDL 817
QY 557 FNIN-----DKRNGPDG---EIOEQHIYNGKEYTSFNDIKQIIDKTLNIIKV 603
Db 818 FNIQNNIPVMSFSLNLSLQFMEIYEKEMVCLYKLDKNDKIKNLEBAKKVSTSV 877
QY 604 KDPARTVTKPEILNKOTGEVSELKPHRVTVTIOGKENSSTIVSEEDFILPVYGELEK 663
Db 878 KTLSSSSMQPLSLTPQDKPEVSANDDTSHSTLNNLSLKLFLNLS-----LGKNKN 928
QY 664 GYQFDGWEISGPEK-----KDAGYVNLISKDTFIKPVFKKIEEKEENKPTF--D 713
Db 929 IVQ-----ELIGOKSENFEKILKUSDFTYNSFTNFVSKADDNLSNDESKRKLLED 984
QY 714 VSKKDNPNQVNSHOLNESHKEDLQREHSQKSDSTKQVATVATVLDKNNISSK-STNNP 771
Db 985 INKLKKTLLQSLFDLYNKKYKLERLFDKKTKYKQWKQIKLTLLEQLESKLSLNNP 1043

RESULT 3
MSPL PLAFW
ID MSPL PLAFW
AC P04913
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMASA) (p195).
GN MSP-1.
OS Plasmodium falciparum (isolate Wellcome).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5848;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=86014355; PubMed=2995820;
RA Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., Riveros-Moreno V.,
RA Nicholls S.C., Hillman Y., Davey L.S., Tizard M.L.V., Schwarz R.T.,
RA Freeman R.R.;
RT "Primary structure of the precursor to the three major surface
RT antigens of Plasmodium falciparum merozoites.";
RL Nature 317:270-273 (1985).
RN [2]
RP REVISIONS.
RA Holder A.A.;
RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X02919; CAA26676.1; -.
DR PIR; A24594; A24594.
DR PIR; S05603; S05603.
DR PDB; ICEJ; 28-MAY-99.
DR InterPro; IPR006209; EGF like.
DR Pfam; PF0008; EGF, 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor; 3D-structure.
FT SIGNAL 1 19
FT CHAIN 20 1639
FT CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 268 268 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 764 764 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 768 768 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 783 783 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 844 844 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 920 920 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 964 964 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1058 1058 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1174 1174 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1445 1445 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1526 1526 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1639 AA; 187618 MW; 2C255B6616C87F6E CRC64;
Query Match 5.4%; Score 218; DB 1; Length 1639;
Best Local Similarity 18.6%; Pred. No. 0.025;
Matches 167; Conservative 164; Mismatches 340; Indels 228; Gaps 39;
QY 23 KDTTCG-VEHHHGHENESIK-----EKSSFTIDRNISTIRDENKDLKGLIKKXFEVD 75
Db 232 KDVGMEDYIKKNTIENIELIEESKTKIDKKNATKBE-----KKLYQAOY 283
QY 76 FTSETGKRMEEYDYKDDRGNIAYDGDGTDEYETEKLEIKSKIYGVLSPKDGHFEIL 135
Db 284 DLSIYNKOLEE-----AHNLISV-----LE---KRIDTLK-----KNIENIKELL 319
QY 136 GKISNVSKNAKYVGNYSKIEIKATKYDFH-----SKTMTFDLYANINDIVDGLAFA 188
Db 320 DKINEIKNPPANGNTPTLLDKNKKIEHEKEIKETAKTIKFNIDSIFTDPL-----373
QY 189 GDMRLFVKNDOKKAEIKIRMEPEKIKETKSEYP-----YVSSYGNV-----IEIGE-GDLSK 239
Db 374 -ELEYLRKN-KNIDISAKVETKESTENEPNGVTYPLSYNDINNALNELNSFGDLI- 430
QY 240 NKPDNLTKMESGKIYSDSEKQOYL--LKDNIIL-----RKGYALKVTTYTPGKTMDL 289

Db 1058 TIISKUKGVIIENVENTENWNTIESSAKEIEALYNELNKKTSLEIYQTSNEVKLOEMKS 1117

Qy 745 KSDSTKDVTA---TVLD---KNNISSKSTNN 770

Db 1118 NADKYIDVSKIFENTVLDTKQSNIVTQHINN 1149

RESULT 5

RPC1 PLAF A

ID RPC1 PLAF A STANDARD; PRT; 2339 AA.

AC P27625;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE DNA-directed RNA polymerase III largest subunit (EC 2.7.7.6).

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_TaxID=5833;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92018020; PubMed=1656254;

RA Li W.B., Bizik D.J., Tanaka M., Gu H., Fox B.A., Inselburg J.;

RT "Characterization of the gene encoding the largest subunit of

RT Plasmodium falciparum RNA polymerase III.";

RL Mol. Biochem. Parasitol. 46:229-240(1991).

CC -! FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION

CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS

CC SUBSTRATES.

CC -! CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +

CC [RNA] (N).

CC -! SUBUNIT: RNA POLYMERASE III CONSISTS OF ABOUT 15 DIFFERENT

CC SUBUNITS. THIS SUBUNIT IS THE LARGEST COMPONENT OF RNA POLYMERASE

CC III.

CC -! SUBCELLULAR LOCATION: Nuclear.

CC -! MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE

CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA

CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE

CC III FOR 5S AND TRNA GENES.

CC -! SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

CC EMBL; M73770; AAA29729.1; -

DR InterPro; IPR007072; RNA_pol A.

DR InterPro; IPR007080; RNA_pol_Rpb1_1.

DR InterPro; IPR007066; RNA_pol_Rpb1_3.

DR InterPro; IPR007083; RNA_pol_Rpb1_4.

DR InterPro; IPR007081; RNA_pol_Rpb1_5.

DR InterPro; IPR006592; RNA_pol_A_N.

DR Pfam; PF04997; RNA_pol_Rpb1_1; 1.

DR Pfam; PF0623; RNA_pol_Rpb1_2; 1.

DR Pfam; PF04983; RNA_pol_Rpb1_3; 1.

DR Pfam; PF05000; RNA_pol_Rpb1_4; 1.

DR Pfam; PF04998; RNA_pol_Rpb1_5; 1.

DR SMART; SM00663; RPOA_N; 1.

DR Transference; DNA-directed RNA polymerase; Transcription; Zinc;

KW Zinc-finger; Nuclear protein. C3H-TYPE (POTENTIAL).

FT ZN FING 88 101

SQ SEQUENCE 2339 AA; 272829 MW; EDD899363086FD48 CRC64;

Query Match

Best Local Similarity 21.6%; Score 214; DB 1; Length 2339;

Matches 191; Conservative 134; Mismatches 306; Indels 254; Gaps 50;

Qy 24 DTGVEHHQOE--NEESIEKESFTIDRNISTIRDF---ENKDLKLIKKKPREVDVDFTS 78

Db 1292 DNTYVEQIEKMKLSKNTKEQSFK-----GTIRDMHEDSEQMNKFITKKAK----FPI 1342

Qy 79 ETGRMEBYDYKDDKGNIIAYDDGTDLEYETEKLDEIKSKIY-----G 122

Db 1343 EK-KKGKHECNDDEIYNTQYD---NIQYNNISNYIKSQNLNTHQVNDLSPKKN 1398

Qy 123 VLSPSKDGH-----FEILGKISNVSKNAKVVYGNYSKIEIKATKYDFHSKMTFPL 174

Db 1399 VILPPKEYHSIFHFVNDYRNVEIKNLMDKKKIFLNNSEKN--VQSKYNRMSKLNKKKI 1456

Qy 175 YANINDIVDG-----LAPAGDMRLFVKNDQ--KKAETIKRMPEK---IKETKSE 219

Db 1457 EI-INNIYRNEKKLNRWKTMDNDNTWSSDDSIIAKKIIKKNKEKKYHPKEKEN 1515

Qy 220 YPVSSYGNVIELGEGDLSKKNPDNLTK-MESGKIYSEKQOYLKKNIIILKRGYALKV 278

Db 1516 PDR-NNYKMITDNNNNNDNNNNNDNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1569

Qy 279 TTYNPGKTDML--EGNGVYSKEDIAKIQKANPNL-----RALSETIYADS-- 322

Db 1570 TNYN---TNIYPNDCNGIYEKETNNNELTTSNMCDKNNDPSEFFNNINENDLYDNKY 1626

Qy 323 -----RNVEDGRSTQSVLMSALDGFNIIRYQVF---TFKMDKGBAIDKDGMLVDSSKL 374

Db 1627 YRQIFKNVIGFVSVEYVESYKQHYILPPEYIIKWTSLELLEYLTIIPTNIFLHTKLSK- 1685

Qy 375 VLFGKD---DKYEYTGEDKFNVEAIKEDGSMFLFDTPKPNVLSMDKNYFNPSSKNKIYVRNP 431

Db 1686 ---KEKPTHQKNTGKMKIYBEIKK--WLFPI--KAINI---YKVFSPKKSIEL----- 1728

Qy 432 EYLRGKISDKGGFNWELRVNVSVDNVLIIYGLDHLIDNTRD-----FNI-----KLVKD 481

Db 1729 -----IKKKDYFNYIIK-NYDISHRYIIH-DYSFINLKQLYLFIFFNYKYFYKYSTP 1779

Qy 482 GDMDWGKDYKANGFPDKVTDMDCNVYLQTVSYDLNAKAVGVHYQFLYDNNKPEVNDP 541

Db 1780 GD-----AVGSSAQSIGEPGTQMTLTKTHFAGVAMWV-TLGV-----PRIKEIINASN 1828

Qy 542 KNTSIEYADGKSVVFNINDKENNGF-----DGEIQEQHIYINGKEYTS-----F 586

Db 1829 SIQTPI---LNIPLVDNNDYNPALMMSKLEKTTIRDCIMYIK-EDYTSRGVFLSVKF 1882

Qy 587 ND-----IKQIDKTLNI-----KI---VVKDFARTTVK--EFLNKOT 621

Db 1893 NEELQKFLNINAYNIKDIILKQSHINKINKINKINHINVKYKLIHLISLKNDEFIF--- 1938

Qy 622 GVSSELKPHRVTVTTIQNGKEMSSITVSEDFILPVYKGELEKGYQFDGWEISGFEGK--K 679

Db 1939 FOMESLKGLDLLLIYGDYKDKIKRCIIKKEDI-----EVTDNDEICD 1980

Qy 680 DAGYVINLSKDTFIKPVFKKIEBK---EENKPTFDVSKKDNFQVNHSQLNESHKED 736

Db 1981 DMDEYNNVSQGT-----ELVERKCSKEENKNAIRVKKEID-----DN 2019

Qy 737 LORBE-----HSQKSDSTKDVATVLDKNNISSKSTNN 770

Db 2020 LEKEENIIYVSEKDSVNLKSEKKDIN---DDNNNDNDNNNN 2060

RESULT 6

COLA_CLOPE

ID COLA_CLOPE STANDARD; PRT; 1104 AA.

AC P43153;

DT 01-NOV-1995 (Rel. 32, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Microbial collagenase precursor (EC 3.4.24.3) (120 kDa collagenase).

GN COLA OR CPE0173.

OS Clostridium perfringens.

OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

OC Clostridium.

OX NCBI_TaxID=1502;

RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 87-113.

RC STRAIN=NCIB 10662 / Type C;
RX MEDLINE=94110220; PubMed=8282691;
RA Matsushita O., Yoshihara K., Katayama S.-I., Minami J., Okabe A.;
RT "Purification and characterization of Clostridium perfringens 120-
RT kilodalton collagenase and nucleotide sequence of the corresponding
RT gene.";
RL J. Bacteriol. 176:149-156(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX MEDLINE=21664373; PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kohara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
RN [3]
RP SEQUENCE OF 1073-1104 FROM N.A.
RC STRAIN=NCIB 10662 / Type C;
RA Matsushita O.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Digestion of native collagen in the triple
CC helical region at Xaa-Gly bonds. With synthetic peptides, a
CC preference is shown for Gly at P3 and P1'; Pro and Ala at P2 and
CC P2'; and hydroxyproline, Ala or Arg at P3'.
CC -!- COFACTOR: Binds 1 zinc ion (By similarity).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M9.
CC -!- SIMILARITY: Contains 1 PKD domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D13791; BAA02941.1; --
CC EMBL; AF003185; BAB79879.1; --
CC EMBL; D50309; BAA08848.1; --
CC PIR; A36866; A36866.
CC MEROPS; M09.002;
CC InterPro; IPR002169; Miccollptase.
CC InterPro; IPR000601; PKD domain.
CC InterPro; IPR006025; Zn_Mrpptcdse.
CC Pfam; PF01752; Peptidase_M9; 1.
CC Pfam; PF00801; PKD; 1.
CC Pfam; PF04151; PPC; 2.
CC PRINTS; PR00931; MICOLLPTASE.
CC SMART; SM00089; PKD; 1.
CC PROSITE; PS00093; PKD; 1.
CC PROSITE; PS00142; ZINC PROTEASE; 1.
CC Hydroxylase; Metalloprotease; Zinc; Zymogen; Signal; Complete proteome.
FT SIGNAL 1 39 POTENTIAL.
FT PROPEP 40 86
FT CHAIN 87 1104 MICROBIAL COLLAGENASE.
FT DOMAIN 774 862 PKD.
FT METAL 502 502 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 503 503 BY SIMILARITY.
FT METAL 506 506 ZINC (CATALYTIC) (BY SIMILARITY).
FT CONFLICT 38 38 L -> F (IN REF. 1).
FT CONFLICT 722 722 I -> M (IN REF. 1).
FT CONFLICT 748 748 G -> E (IN REF. 1).
FT CONFLICT 945 945 V -> E (IN REF. 1).
FT CONFLICT 970 970 T -> A (IN REF. 1).
FT CONFLICT 987 987 A -> E (IN REF. 1).
FT CONFLICT 1098 1098 I -> T (IN REF. 1 AND 3).
SQ SEQUENCE 1104 AA; 125935 MW; F4B7377194ED021C CRC64;

Query Match 5.2%; Score 209.5; DB 1; Length 1104;
Best Local Similarity 20.0%; Pred. No. 0.036;
Matches 194; Conservative 136; Mismatches 345; Indels 295; Gaps 50;

QY 6 AESKFNKLGNGKESGLKDDTTGVEHHQENBESIKESSTFTIDRNISTIRDFENK----- 60
Db 40 ADKVKENSNGQ---ITREINADQISKTELNNEVATDNNRPLGSPISAPSRARNKIYTPD 96
QY 61 -----DLKKLIK-KFREVDD-FTSETGK-----RMEFYDYKYDDKGNIIAYDD 102
Db 97 ELNRMYSDLVELIKTISYENVDPDLFNFNDGSYTFFSNRDRVQAIYIGLEDSGRITYADD 156
QY 103 GTDL-----EYETEKLDEIKSIYK-----VLSPSKDGHF 132
Db 157 DXGIFLVEFLRAGYVLYGYNKQLSYLNTPLQKNECLPAMKAIQNSNFRJGTKAQDGVV 216
QY 133 ELIGK- ISNVSKNAKY-----YGNNY-----KSIKATKYDF 165
Db 217 EALGRLGNASADPEVINNCIVVLSDFKDNIDKYSNYSKGNVFNLMKGIYDYNISVIY 276
QY 166 HSK---TMTFDLYANI-----ND-----IVDGLAFAGDMRLFKVDNQD 200
Db 277 NTKGYDAKNTFYNRIDPYMERLESCTIGDKLNDNNAWLVNVALYVYTCRMGKFREDPSI 336
QY 201 KKAIEKIRMEPEKIKETKSEYPYVS-----SYGNVIELGEGDLSKNKPDNLTKMESGKIYSD 256
Db 337 SQRALERAM-----KEYPILSYQYIEAANDLDLNF--CKNSSGN--DIDFNKIKAD 384
QY 257 SEKQVLLKDNIIILRKGYALKVTTYNPGKTDMLGNGVYSKEDIKIAKQANPNLRALSET 316
Db 385 A-----REKYLPTYTFDDGKFVVKAGDKV-TEKIKELYASKEVKAQFNR 430
QY 317 TIYADSRNVEDGRSTQSVLMSALDGFNIRYQVTFKMD--KGEAIDKQGNLVTDSKL 374
Db 431 VYQND- KALEGNPDILTV-----VIYNSPEYKLNRIINGFSTONGYIENIGTF 482
QY 375 VLFQKDDKE--YTGEDKNVE-----AIKEDGSMFLDITKPVNL--- 411
Db 483 FTYERTPEESIYTELBEFRHEFTHYLQGVVYVPGMWGQGEFYQEGVLTYEESGTAEFFAG 542
QY 412 SMDKNVFNPSKS---NKIYVRNPEFYLGRKISDKGGFNWELRVNESVDNLIYGLDHD 468
Db 543 STRTDGIRKPSVTOGLAYDRNRMSLYGLHAKYG-SWDFYNYGFALSNY-----MYNN 596
QY 469 NTRDFNIKLN-VKGDGIMDMGMDKYKAN-----GFPDKVTD-MDGNVYLQGYSDLNKA 521
Db 597 NMGFMENKMTNLIKNDVS--GYKDIASMSDYGLNDKYQDYMS---LLNNIDNLDVPL 651
QY 522 VGVHY-----QFLYDNVKEVNI-DPKGNTSIEYADGKSVVFNINDKNNNGFDGRI 571
Db 652 VSDEYVNGHEAKDINEITNDIKEVSNIKDLSNVE-----KSQFTTYDMRGV----- 699
QY 572 QEQHIYINGKEYTSFNDIKQIIDKTLNLIKIVVKDFAR-----NTTVKEFILNKDGTGE-- 623
Db 700 -----YVGRSOGCEENDWK---DMNSKLNLDILKELSKSWNGYKTVTAYFVNHKVDGNGN 751
QY 624 -VSELKPHRV-----TVTIQNGKEMSTIVSEEDFLPVYKBLEKGYQPDGWEISGFEGK 678
Db 752 YVVDVVFHGMNTDTNTDVHVNKEPKAVIKSDSVI-----VEBEINPDGTESKDEDEGE 804
QY 679 -----KDAYVINLSK-----DTFIKPVFKLEEKKER- 706
Db 805 IKAYEWDFGDGBKSNEAKATHKYNKGTGEYEVKLTVDNNGNGINTESKKT-KVEDKPEVE 863
QY 707 --ENKPTFDVSKKKNQPNVHNSQLNESHK--KEDIQREHSQKSDSTKDVATATVLDKNNI 762
Db 864 INESEPNDFEKA-----NQIAKSNMLVKGTLSSEEDYSK-----YFPDVAKKGNV 909
QY 763 SSKSTNNPN 772
Db 910 --KITLNNLN 917

RESULT 7
SPOF_SCHPO
ID SPOF SCHPO STANDARD; PRT; 1957 AA.
AC Q10411; Q9USE9;

DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Sporulation-specific protein 15.
 GN SPO15 OR SPAC1F3.06C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A., FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.
 RX MEDLINE=20107136; PubMed=10639340;
 RA Ikenoto S., Nakamura T., Kubo M., Shimoda C.;
 RT "S. pombe sporulation-specific coiled-coil protein Spo15p is localized
 RT to the spindle pole body and essential for its modification";
 RL J. Cell Sci. 113:545-554(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volkart G., Aert R., Robben J., Crymonprez B.,
 RA Weljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Babor C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Fumelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe";
 RL Nature 415:871-880(2002).
 RN [3]
 RP SEQUENCE OF 705-871 FROM N.A.
 RC STRAIN=968 h90;
 RX MEDLINE=20223868; PubMed=10759889;
 RA Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,
 RA Hirooka Y.;
 RT "Large-scale screening of intracellular protein localization in living
 RT fission yeast cells by the use of a GFP-fusion genomic DNA library";
 RL Genes Cells 5:169-190(2000).
 CC -!- FUNCTION: Has a role in the initiation of spore membrane
 CC formation.
 CC -!- SUBUNIT: Monomer.
 CC -!- SUBCELLULAR LOCATION: Spindle pole body.
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; 270690; CA94624.1; --
 CC EMBL; AB027811; BAA87115.1; --
 CC PIR; T38077; T38077.
 CC GenBank_Spombe; SPAC1F3.06c; --

KW Sporulation; Coiled coil.
 FT DOMAIN 199 785 COILED COIL (POTENTIAL).
 FT DOMAIN 804 1235 COILED COIL (POTENTIAL).
 FT DOMAIN 1320 1471 COILED COIL (POTENTIAL).
 FT DOMAIN 1481 1723 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1957 AA; 222785 MW; 3F480CA06171D9DA CRC64;
 Query Match 4.8%; Score 194; DB 1; Length 1957;
 Best Local Similarity 20.5%; Pred. No. 0.33;
 Matches 174; Conservative 147; Mismatches 290; Indels 238; Gaps 44;
 QY 4 ETAESEKFNGLNGKESGLKQDTTGVHHQENEE-----SIKESSFTIDRNISTI 54
 DB 937 KIEESKSDLG-----KKLTARQEEISLNKEENSSQAITSVKSKUDELTSKSKLE 989
 QY 55 RFENKDLKKIJKKIFREVDFTSETGKRMEEYDYKDD---KGMIIAYDDGDTLEETE 111
 DB 990 ADIEH-----LNKVSEVE---VERNALLASNERLMDDLKNGENIA-SLQTEIEKKRA 1039
 QY 112 KLDEITSKIYGVLSPSKDGHFELGKISNVSKAKVYVYNNYKSIIEIKATKYDFHSKWT 171
 DB 1040 ENDDLQSK-----LSVVSSEYENLLLISSQTNKSLDKTNQLKYIEK--- 1081
 QY 172 FPLYANINDVDGLAFAGDMRLFVKDNDQKAEIKRMPEKIKETKSEYPYSSVCGNVIE 231
 DB 1082 -----NVQKLLD-----EKOQRNVELE-----ELTSKYGKGL-- 1107
 QY 232 LGEGLSKNKPDLNLTWESGKIYSDSEKQYLLKDNII--LRKGVALKVTVTPNGKIDML 289
 DB 1108 ---GEENAIKDELLALRK-----KSKQKQHLDFANFVDLKE-----KSDAL 1146
 QY 290 EGNVYSKEDIKIQKPNPLALSE-----TTIYADSRNVEDGRSTOSVLMSALDGNF 343
 DB 1147 EQLTNEKNELIVSLQSSNNNEALVEERSDLANRLSDMKKSLSDSNVSVIRSDL--- 1202
 QY 344 IIRYQVFFKMDKGEADKQGNLVTDSKLVLFKDD--KYTG-EDKFNVEAIKEDGS 400
 DB 1203 -----VRVNDLDTLKKDKDSLSTQYSEVQCDRDDLLSLKGEESFNKYAV---S 1250
 QY 401 MLFIDTK-----PVNLMDKNYF---NPSKNKIVVRNPEFYLRGKISDKGFMWELRV 451
 DB 1251 LRELCTKEIDVPVSEIILDDNFVFNAGNFSELSRTVLSLENYL-----DAFN-QVNF 1302
 QY 452 NESVVDNYLIYGLHIDNTRDFNIKLN-----VKDGDIMDMKMDYKANGFPDKVTOM 504
 DB 1303 KMWELDNRLTTTDAEFTKVVADEKLQHEHDDWLIQGD-LEKALKDSEKN-FLRKEAEM 1360
 QY 505 DGNVY-LOTGYSDLNKAVGVHYQFLYDNV-----KPEVNIDPKGNTSIYADG-----KS 554
 DB 1361 TENIHSLEEGKEE-TKGEIAELSLRLEDNQLATNKLKNQDLHL-NQEIHLKEDVLKES 1418
 QY 555 VVFNINDKRNNGFDGEIOBOHIYINGKEYTSFNDIKQIIDKT--LNKIVVKDFARNTVV 612
 DB 1419 LIISLEESLSN-----QROKESLLDAK-----NELEHLDLDTSRKNSSLMKESINSLSL 1469
 QY 613 --KEFILNKDTGEVSEL-KPHRVTVT-----IQNGKE---MSSTIVSEEDFILPVYK 658
 DB 1470 DDKSFELASAVEKLGALQKLHSELSLMENIKSQLQEAKEKIQVDESTIQELDHEITASK 1529
 QY 659 GELEKGYQFDGWEISGFGKKDAGVIVNLSKDTFTKPVFKKIE----- 702
 DB 1530 -----NNYEGKLN-----DKDSIIRDLSENTEIQLNLLAEBKSAVKRLS 1568
 QY 703 -KKEEE---NKPTFDVS--KKDNQPVNHSQLNESHKEDLQREHSQKSDTK--DVT 753
 DB 1569 TEKESIEQLQFNRLADLEYHKSQVSELSGRSKLKLASTTEELQLAENERLSLUTRMLDQ 1628
 QY 754 ATVLDKNNI 762
 DB 1629 NOVKDLSNI 1637

RESULT 8

RA Kendrick K.E.;
 RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-8 FROM N.A.
 RA Bai Y., Symington L.S.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC 1- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
 CC COMPLEX.
 CC 1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
 CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
 CC ER AND THE GOLGI COMPLEX.
 CC 1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
 CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
 CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
 CC 1- SIMILARITY: BELONGS TO THE VDP/USOL/YBL047C FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X54378; CAA38253.1; -
 CC EMBL; L03188; AAB00143.1; -
 CC EMBL; U53668; AAB66659.1; -
 CC SGD; S0002216; USOL.
 CC InterPro; IPR002017; Spectrin.
 CC InterPro; IPR006955; Usol_p115_C.
 CC InterPro; IPR006953; Usol_p115_head.
 CC Pfam; PF04871; Usol_p115_C; 1.
 CC Pfam; PF04869; Usol_p115_head; 1.
 CC Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
 KW DOMAIN 1 724 GLOBULAR HEAD.
 FT DOMAIN 725 1790 COILED COIL (POTENTIAL).
 FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
 FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
 FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).
 FT CONFLICT 847 847 G -> E (IN REF. 2).
 FT CONFLICT 924 924 E -> K (IN REF. 2).
 FT CONFLICT 1253 1253 V -> I (IN REF. 2).
 FT CONFLICT 1319 1319 I -> V (IN REF. 2).
 FT CONFLICT 1461 1461 N -> S (IN REF. 2).
 FT CONFLICT 1581 1581 G -> S (IN REF. 2).
 FT CONFLICT 1600 1600 I -> V (IN REF. 2).
 FT CONFLICT 1661 1661 R -> S (IN REF. 2).
 FT CONFLICT 1772 1772 D -> DEEDDEE (IN REF. 2).
 SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;
 Query Match 4.6%; Score 186.5; DB 1; Length 1790;
 Best Local Similarity 18.1%; Pred. No. 0.63;
 Matches 156; Conservative 158; Mismatches 277; Indels 269; Gaps 41;
 QY 1 KLGEIAESK-----FKNLNGKEGSIKKDTTGVHHHQBESIK-----EKSFF 45
 DB 1076 KISELTKTBREAEALAAVANKNELETKLTSEKALK-EVENEHLEKBEKIQLEKAT 1134
 QY 46 TIDRNISTIR-----DFENKDKLKLKKFREVDFTSETGKRMEEYDYKYDDKGNIIA 99
 DB 1135 ETQQLNSLRANLESLEKEHDLAAQLK-----YEQIANKEQYNEEISQLN---- 1183
 QY 100 YDGTDLVETE-----KLDRISKIYGVLSPS-----KDGHPFELG-----KISNVSK 143
 DB 1184 -DEITSTQGENESIKKNDLEGEVAMKSTSEQSNLKKSEIDALNLQIKELKKKNETN 1242
 QY 144 NAKVYGNVVKSTIEIKATKY-----DFHSKTMFTDLVANIINDIVDGLAFAGDMELFVK 196
 DB 1243 EASLL--ESIKSVESSTVIKELQDCNFKEK-----EVSEEDKUKASDKNSKYL 1292
 QY 197 DNDOKKAEIKIRPEKIKETKTSYPVSSYGVNIELGEGDLS-----KNKPDNLIT 246
 DB 1293 ELQKSEKIKEELDAXTELKIQLEKILTNSLAKAKESSELSRLKKTTSBEERKNAEQLE 1352

QY 247 KMSG-KIYSDS-EKQOYLKLD-NILRKGYALKVTTYNPGKTDMLGNGVYSKE----- 298
 DB 1353 KUKNEIQKNQAFERKLLNEGSSSTITOEYSEKINTLEDELIRIQENELKAKELDNT 1412
 QY 299 -DIAKIQANPNLRALSETTIVADSRNVEDGRSTOSVLSALDGFNIIRYQVFTFMNDK 357
 DB 1413 SELEKVSLSNDELLEEKQNTI---KSLQD-----EILSKDKITRNDK 1453
 QY 358 GRAIDKDGMLVTDSSKLVLFQKDDKEYTGEDKFNVEAIKEDGSMFLPIDTQPNLSMDKAY 417
 DB 1454 LLSIERDNKRDLSESLKEQLRAAQESKAKVEE--GLKKLEESS-----KEKAELEKSKEM 1506
 QY 418 FNPSSNKIYVNPREFYLRGKLSKGGFNWELRVNESVVDNLYLYGDLHIDNTR----DF 473
 DB 1507 MKKLEST---ISNETELKSSME-----TIRKSEKLEQSKSAEEDIKULQHEKSDL 1556
 QY 474 NIKLVNKGDMIDWGMKDYKANGFPDKVTDMDGNVYLOTGYSDLNAKAVGVHYQFLYDNV 533
 DB 1557 ISRINESEKI-----EEL 1570
 QY 534 KPEVNDPKGNTSIEYADGKSVVFNINDKRNNGFGHEIOEQHIYINGKEYT-----SFNDI 589
 DB 1571 KSKLRIEAKSGSELE-----TVKQELNNA-----QEXIRINAENTVLKSKLEDI 1615
 QY 590 -KQIIDKTLNIIKIVKDFARNTTVKEFILNKDTGEVSELKPHRVTVTIQNGKMSST--- 645
 DB 1616 ERELKDKQAEIK-----SNOEKELL-----TSRLKELE-----QELDSTQOK 1653
 QY 646 -IVSEEDFILPVYKGELEKGYQFD-----GWEISGFEGKKDAGYVIN 686
 DB 1654 AQKSEERRAEVRFQVEKS-QLDKAMLETKYNDLVNKEQAW-----KRD----- 1699
 QY 687 LSKOTFIKPVFKIEKEEENKPTFDVSKKDKNQVNSHLSQHNESHKEDLQREHSQKS 746
 DB 1700 --EDT---VKTTDSQREIEK---LAKELDLNKAENSKLKEAN-----EDRSEID 1742
 QY 747 DSTKDVATVLDKNISKS 766
 DB 1743 DLM--LLVTDLDEKNAKYRS 1760
 RESULT 11
 ID GIAN HUMAN STANDARD; PRT; 3259 AA.
 AC Q14789; Q14398;
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE Golgi autoantigen, golgin subfamily B member 1 (Giantin) (Macrogolgin)
 DE (Golgi complex-associated protein, 372-kDa) (GCP372).
 DE GOLGB1.
 GN
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94187728; PubMed=7511208;
 RA Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,
 RA Renz M.;
 RT "Molecular genetic analyses of a 376-kilodalton Golgi complex membrane
 RT protein (giantin).";
 RL Mol. Cell. Biol. 14:2564-2576(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94257116; PubMed=8198703;
 RA Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,
 RA Renz M.;
 RT "Macrogolgin -- a new 376 kD Golgi complex outer membrane protein as
 RT target of antibodies in patients with rheumatic diseases and HIV
 RT infections.";
 RL J. Autoimmun. 7:67-91(1994).

[3]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=95100974; PubMed=7802676;
 RA Shoda M., Misumi Y., Fujiwara T., Nishioka M., Ikehara Y.;
 RT "Molecular cloning and sequence analysis of a human 372-kDa protein
 RT localized in the Golgi complex.";
 RL Biochem. Biophys. Res. Commun. 205:1399-1408(1994).
 CC !- FUNCTION: May participate in forming intercisternal cross-bridges
 CC of the Golgi complex.
 CC !- SUBUNIT: Homodimer; disulfide-linked.
 CC !- SUBCELLULAR LOCATION: Membrane-associated protein. Golgi.
 CC !- DISEASE: Antigen in chronic rheumatoid arthritis and in the
 CC autoimmune disease Sjogren's syndrome.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X75304; CAA53052.1; -;
 DR EMBL; D25542; BAA05025.1; -;
 DR FIR; A56533; A56533.
 DR FIR; I52300; I52300.
 DR Genew; HGNC:4429; GOLGB1.
 DR MIM; 602500; -;
 DR GO; GO:000139; C:Golgi membrane; TAS.
 DR GO; GO:0005795; C:Golgi stack; TAS.
 DR GO; GO:0016021; C:integral to membrane; TAS.
 DR GO; GO:0007030; P:Golgi organization and biogenesis; TAS.
 KW Golgi stack; Antigen; Coiled coil; Transmembrane.
 FT DOMAIN 1 3235
 FT TRANSMEM 3236 3256
 FT DOMAIN 3257 3259
 FT DOMAIN 48 593
 FT DOMAIN 677 1028
 FT DOMAIN 1062 1245
 FT DOMAIN 1301 1779
 FT DOMAIN 1828 3195
 FT DOMAIN 2420 2423
 FT DOMAIN 2993 2996
 FT CONFLICT 1 39
 FT CONFLICT 215 215 A -> AQLSSM (IN REF. 3).
 FT CONFLICT 1765 1765 D -> G (IN REF. 3).
 FT CONFLICT 2950 2950 H -> D (IN REF. 3).
 SQ SEQUENCE 3259 AA; 376075 MW; 60376A20D8A178DD CRC64;
 Query Match 4.6%; Score 185; DB 1; Length 3259;
 Best Local Similarity 20.0%; Pred. No. 1.5;
 Matches 170; Conservative 150; Mismatches 338; Indels 194; Gaps 39;
 QY 21 LKQDTGVHHQENEE--SIKESSFTDRISTIRDFENKDLKLIK--PREVD 75
 DB LKRIAGLBEKQKNEKFNKTLNLSQIST-KDGLKMLQEVTKMNLNQIQIE 1903
 QY 76 FTSETGKREEDYKYDD-----KGNIIAY-DGDTLVEYETELDEIKSIY 122
 DB ELRVTVKLKEETAEKDDLEERLMQLAELNSIGNYCDVTDQAIKNELLESEMKNLKK 1963
 QY 123 VLSPSKDGHFEILGKISVSNKAKVY-----GN-----NYKSTIEIKA 160
 DB CVSELEEEKQQLVKEKTKVESEIRKEYLEKIQAQKEPGNKSHAKELQELLKEQEVQK 2023
 QY 161 TKYD---FHSKWTTFDYANINDVDGLAFAGDMRLFVKNDQKKAIE-KIRPEKIKET 216
 DB LQKDCIRYQEKISALERTVKALE-----FVQTESQKOLEITKENLAQAVEHR 2070
 QY 217 KSEYPTVSSYGNVIELGEGDLSKGNKDNLTVMESGKIYSDSEKQVLLKDNILRK---- 272
 DB KKAQAEALASFKVLLDPTQSEAAARVLADNLKIKELQSNKESVSKQMKQKDELLERLEQA 2130

QY 273 -GYALKVTYTPPKTMDLEGNVYKSKEDIATKQKPNLRALSETTIYADSRNVEGRST 331
 DB 2131 EEKHLKEKKMOEKLDALRREKVHLEETIGI-----VTLNKKKEVQVQ---- 2175
 QY 332 QSVLSALDGFNIIRVQVFTFQNDKGEAIDKGNLVTDSSKLVLFKGDQKVEYTGDKFN 391
 DB 2176 ---LQENLDS-TVTQLAAFTKMS-----SLQDRDRVIDEAK-----KWERFSDAIQSK 2222
 QY 392 VBAI--KEDG-SMLFIDTKPNVLSMDKNVFNPSK---SNKIY--VRNPFYLRGKISDK- 442
 DB 2223 EBEIRLKEDNCVSLQDLRQMSIHMEELKINISRLSHDKQIWESKAQTEVQLQKQVCDTL 2282
 QY 443 GGFNWLVRNESVVDNYLIYDGLHDINTDFNKLNVKGDIDMDGMKDYKANGFPDKVT 502
 DB 2283 QGENKELL-----SOLEETR-----HLVHSSQNEKLAKESELKSLKDQUT 2322
 QY 503 DMD-----GNV--YLTGYSDL-NAKAVGVHYQFLYDNVKNPEV----- 537
 DB 2323 DLSNLEKCKEKGKMLGIIROEADIQNSK-----FSYQLETDLQASRELTSLRHE 2375
 QY 538 NIDPKGNTSIEYADGK--SWFNIINDKRNNGFDGEIQEHIYINGKEYTSF---NDIKQI 592
 DB 2376 EINMKEQKIISLSCKEBAIQVAIAELRQO-HDKEIKELLENLSQEBEENIVLEENKKA 2434
 QY 593 IDKTLNLIKIVKDFARNTTVKFEILNKDTGEVSELK--PHRVTVTIQNGKMSSTIVSEE 650
 DB 2435 VDKTNQLMETLTKIKENIQKRAQLDSFVKSMSSLQNDRDRIVDGYDQLEERHLSILSK 2494
 QY 651 DFILPVYKGELEKGYQDFGWEISGPEG-----KKDAGVYVNLKSDT 691
 DB 2495 DLIQEAANAENKLEK-----EIRGURSHMDDLNSENKALDABLIQYREDNQVITI-KUS 2549
 QY 692 FIKPVFK-KIEEKEENKPTFDVSKKDKNDPQVNHSQLNES-----HRKEDLQREHSQK 745
 DB 2550 QOKLLEVLQOKNLEKNKYKLEKLEKSEBAN-EDLRRSFNALQEEKQDLKSLETSLK 2608
 QY 746 ---SDSTKDVTA 754
 DB 2609 VSISQLTRQVTA 2620
 RESULT 12
 ID MLPI_YEAST STANDARD; PRT; 1875 AA.
 AC Q02455;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin-like protein MLPI.
 GN MLPI OR YKR095W OR YKR415.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c;
 RX MEDLINE=93247549; PubMed=8483450;
 RA Koelling R., Nguyen T., Chen E.Y., Borstein D.;
 RT "A new yeast gene with a myosin-like heptad repeat structure.";
 RL Mol. Gen. Genet. 237:359-369(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94205265; PubMed=8154186;
 RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
 RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.D.;
 RT "The complete sequence of a 15,820 bp segment of Saccharomyces
 RT cerevisiae chromosome XI contains the UBI2 and MPL1 genes and three
 RT new open reading frames.";
 RL Yeast 9:1349-1354(1993).
 CC !- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
 CC REPAIR.
 CC !- SIMILARITY: SOME, TO THE TPR ONCOGENE.

CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC HEXAGONAL S-LAYER.
CC -!- SIMILARITY: Contains 3 s-layer homology (SLH) domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M15364; AAA87321.1; --
CC EMBL: M14238; AAA22372.1; --
CC EMBL: M19115; AAA22760.1; --
CC EMBL: M31828; AAA22619.1; --
CC PIR: A28555; A28555.
CC InterPro: IPR001119; SLH.
CC Pfam: PF00395; SLH; 2.
CC PROSITE: PS01072; SLH DOMAIN; 2.
CC Cell wall, S-layer; Signal; Repeat.
CC SIGNAL 1 23
CC CHAIN 24 1053 MIDDLE CELL WALL PROTEIN.
CC DOMAIN 29 92 SLH 1.
CC DOMAIN 93 143 SLH 2.
CC DOMAIN 144 203 SLH 3.
CC SEQUENCE 1053 AA; 117146 MW; DB4213188D9D5E4F CRC64;
SQ
Query Match 4.6%; Score 183.5; DB 1; Length 1053;
Best Local Similarity 20.2%; Pred. No. 0.45;
Matches 194; Conservative 143; Mismatches 326; Indels 297; Gaps 52;
QY 10 FKNLNGKESLKKDTT-GVEHHQHNEESIKESFTIDRNISTIRDFE-----NK 60
DB 183 FKNLNDALVDMQVEFGDIRHEITKTLTKYLKVV-----TVRDMWAQEAAGNDSE 236
QY 61 DLK-----KLIKKKFREV-----DDFTSETGKMEYDYKY-----DDKG 95
DB 237 DPLPLTVNPAIGLGIKANEVTLNGKDGAGINTTYKVADGINANDFGQHVQVWIKDDKE 296
QY 96 NIIAYVDG-TLEYETEKLDE--IKSKIYGLVSPKDGHFELIGKISNVSKAKVYVGN 152
DB 297 DVIVMEGTDQVIMDRVGEFTLKGT-----FEDPKDLSN-SDLADL----- 339
QY 153 YKSIETKATK--YDFHSKWTTFDLVANIINDIVDGL-----AFAGDMRLFVKNDQ 200
DB 340 --KLELDASEKSYRKNKTVNFTFRFNDPVDGLKEIKNADGCGFTFGAKV-VLDNN 396
QY 201 KKAETIKRMPKIKETKSEYVPVSSYGVNVELGE-----GDLS----- 238
DB 397 EIAYIHVIDDQSMNKEDEGVYGVSEVISKIDTKKKITNRDNDKFNLDGKEEGDFLVF 456
QY 239 -KNKPDNLTKMESGKIYS-----DSEKQOYLKKNILKGVALKVTVYVNPCKTMDL 291
DB 457 LNKPAKFDLKEGWMYVYVYADGDDKLLVFATDTV--EGKVDKVSRRNNVRLTIG 514
QY 292 NGVY-----SKEDIAIKQANPNL-RALSETTI--YADS-----RNVE-----DG 328
DB 515 DKTYRVEGATSDGDKNDQVQIDKDHDLVDSDDTEVKYLDASGRVPHIETKDAIDD 574
QY 329 RTQSVMALDGFNIIRVQVFTFK-MNDKGRAI-----DKDGNLVTSSKLVLV 377
DB 575 RK-QRAIVTRSATNTSK-DTWDFRVLTKQKEITVSLKAKNIYDPDGKNFSDRNK----- 628
QY 378 GKDDKE---YTGEDK-----FNVEAIKEDGSMFLFDTKPNLSMDKN----- 416
DB 629 NQDDLEDILVPSKDDTLLELVTLADGPGKVEFL--KPVKVEQESGKAWDDLADDD 686
QY 417 -----YFN-----PSKNKIYVRNPEFLYRIGISKDGKGNFWELRVNES----- 454
DB 687 MVGDVEVTKTAVFNMTGKLESSKKELKNAKTAFAKVDADENDLSVITYVNDKDEVEA 746
QY 455 --VVDNYLYGLDHLIDNTRDF-----NIKLVNVDGD-IMDWGMKDYKANGFPDKVTDM 504

DB 747 IFVVEGDGLTGAHYGVQVDFGRKGGKOTIRVWEKDGDKVBE---KEYLDDGGDDDKDE 803
QY 505 DGNVYLTQTSYDLNKAAGVHYQFLYDNNVKNPEVNDPKGNTSIEYADGKSVVFNINDKN 564
DB 804 D-----IRNDFFIAFTVDSNDEVVDDVEVNNKAKGMLA-EVTDKGMKDANIDK 856
QY 565 NGFDGEIQSQHIIYINGKEYTSFNDIKIIDKTLNLIKIVKDFARNVTVEKFLNKDTGEV 624
DB 857 VGLVSDVRKDTI-----TYKDADNKK-----KASIKSATV-YFDLYDDFGEA 898
QY 625 SELKPHRVTVTTIQNGKMSSTIVSSEDFILPVVKGLEKGYQFDGWEISGFEGKDGAGV 684
DB 899 DG-----VNGDYVVMIDSGDI-SGTKYD-----YV 923
QY 685 INLSKDTFIKPVFKKIEE-----KKEENKP-----TFDVSKKKONPQVNH 726
DB 924 LIVSDAKTVRK--DKLEDDAEAFKQEPSEKPDPTKWDALPSKVGKFTSAGPVKLYRA 981
QY 727 --QLNESHKEDLQREE--HSQKSDST-----KQVTV-----LQKNISSTTNNPK 773
DB 982 TVELNSKVKAEVDVAIEFYFNGKVEPSLLNFKDGVTITGNTEDKVTSSKIKVTNKGK 1041
RESULT 14
WAPA_BACSU
ID WAPA_BACSU STANDARD; PRT; 2334 AA.
AC Q07833;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Wall-associated protein precursor.
GN WAPA OR N17G.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=93302506; PubMed=8316082;
RA Foster S.J.;
RT "Molecular analysis of three major wall-associated proteins of
RT Bacillus subtilis 168: evidence for processing of the product of a
RT gene encoding a 258 kDa precursor two-domain ligand-binding
RT protein.";
RL Mol. Microbiol. 8:299-310(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSC1A1;
RX MEDLINE=95219088; PubMed=7704263;
RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
RT genome containing the hut and wapa loci.";
RL Microbiology 141:337-343(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSC1A1;
RX MEDLINE=97124196; PubMed=8969509;
RA Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,
RA Miwa Y., Fujita Y.;
RT "Sequencing of a 65 kb region of the Bacillus subtilis genome
RT containing the lic and cel loci, and creation of a 177 kb contig
RT covering the gnt-sacX region.";
RL Microbiology 142:3113-3123(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,

RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koester P., Konigstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauei C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadaie Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Takatoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weltzenegger T.,
 RA Winters P., Wipat A., Yanamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis";
 RL Nature 390:249-256(1997).
 CC -!- FUNCTION: STILL UNKNOWN. NOT INVOLVED IN CELL MEMBRANE METABOLISM,
 CC MOTILITY, SECRETION OR DIFFERENTIATION.
 CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED
 CC INTO THE MEDIUM.
 CC -!- DOMAIN: HAS TWO LIGAND-BINDING DOMAINS; THE N-TERMINUS, HAS THREE
 CC 101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE
 CC C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED
 CC MOTIF REPEATED 31 TIMES.
 CC -!- SIMILARITY: THE REPEATED MOTIF IN THE C-TERMINUS HAS SOME
 CC SIMILARITY TO THE REPEAT IN E. COLI RHS GROUP OF PROTEINS (RHS-A-D).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL; L05634; AAA22883.1; -;
 CC EMBL; D31856; BAA06656.1; -;
 CC EMBL; D29985; BAA06260.1; -;
 CC EMBL; D83026; BAA11683.1; -;
 CC EMBL; Z99124; CAB15959.1; -;
 CC FIR; S32920; S32920.
 CC Subtilisin; BG10797; wapa.
 CC InterPro; IPR003305; CBM_Cenc.
 CC InterPro; IPR006530; YD_
 CC Pfam; PF02018; CBM_4_9; 1;
 CC TIGRFAMs; TIGR01643; YD_repeat_2x; 17.
 KW Cell wall; Repeat; Signal; Complete proteome.
 FT SIGNAL 1 28 OR 32 (POTENTIAL).
 FT CHAIN 29 2334 WALL-ASSOCIATED PROTEIN.
 FT DOMAIN 504 869 3 X 101 AA APPROXIMATE TANDEM REPEATS.
 FT REPEAT 504 605 1-1.
 FT REPEAT 636 736 1-2.
 FT REPEAT 769 869 1-3.
 FT DOMAIN 1021 2139 31 X 21 AA APPROXIMATE TANDEM REPEATS OF
 FT X(4)-G-X(4)-[YF]-X-D-X(2)-G-X(4).
 FT REPEAT 1021 1040 2-1.
 FT REPEAT 1042 1061 2-2.
 FT REPEAT 1063 1082 2-3.
 FT REPEAT 1083 1102 2-4.
 FT REPEAT 1109 1128 2-5.
 FT REPEAT 1129 1148 2-6.
 FT REPEAT 1150 1169 2-7.
 FT REPEAT 1174 1193 2-8.

FT REPEAT 1199 1218 2-9.
 FT REPEAT 1219 1238 2-10.
 FT REPEAT 1667 1686 2-11.
 FT REPEAT 1690 1709 2-12.
 FT REPEAT 1711 1730 2-13.
 FT REPEAT 1732 1751 2-14.
 FT REPEAT 1753 1772 2-15.
 FT REPEAT 1795 1814 2-16.
 FT REPEAT 1820 1839 2-17.
 FT REPEAT 1840 1859 2-18.
 FT REPEAT 1861 1880 2-19.
 FT REPEAT 1887 1906 2-20.
 FT REPEAT 1908 1927 2-21.
 FT REPEAT 1929 1948 2-22.
 FT REPEAT 1969 1982 2-23.
 FT REPEAT 1983 2002 2-24 (APPROXIMATE).
 FT REPEAT 2008 2027 2-25.
 FT REPEAT 2028 2047 2-26.
 FT REPEAT 2051 2070 2-27.
 FT REPEAT 2071 2090 2-28.
 FT REPEAT 2093 2112 2-29.
 FT REPEAT 2120 2139 2-30.
 FT REPEAT 2139 2158 2-31.
 SQ SEQUENCE 2334 AA; 258329 MW; B75138CCD278BAA3 CRC64;
 Query Match 4.5%; Score 183; DB 1; Length 2334;
 Best Local Similarity 20.9%; Pred. No. 1.2; Indels 320; Gaps 45;
 Matches 195; Conservative 99; Mismatches 317;
 QY 51 ISTIRDFENKDKLKKLKKF-----REVDFTTSETGKMEBYD-----88
 DB 1168 IQVIDDAEG--LKIITNTKYEGNNVEDVDNDVGTKATESYQYDKDGNVTSVKDAYGT 1225
 QY 89 --YKY-----DDKNI--IAYDDGTDEYETEKLDEIKSKI---YG-----V 123
 DB 1226 ETEYNNKNDVTMKDKTEGNTVDIAY--DGLDAVSETDQSGKSSAAVYDKYGNQIOSSKD 1284
 QY 124 LSPS-----KDGHEI-----LGKISNVSKNAKVVYGNVYKSTIEI-----KAT 161
 DB 1285 LSASNILKDGSEFAQKSGWNLTASKDRKKIIVADKSGVLGSG--KALEVLSQSTSGT 1342
 QY 162 KYDFHSKWTFTDLYANI-----NDIVDGLAFAGDMRLFYKNDQKAEIKIRMEK 212
 DB 1343 DHGYSSATQVELEPNTTYTSLGKIKTLAKSRAY---FNIDLRDKQK-----R 1389
 QY 213 IKETSEYYPVSSYGNVIELGGLSLKNPDLNLTWESCKI-----YSDSEKQYLLK 265
 DB 1390 IQWIEHNEYSALA-----GKNDWTKRQITFTTPANAGKAVVYMEVDHKDKGKAWF 1441
 QY 266 DNIILRKGYALKVTVYVPGKTDN-----LEGNGVYSKEDIKIAIKANPNLRALSET 316
 DB 1442 DEVQLEKGEV--SSSNPVPQNSFTSATENNVNVSAGASVDSSEGF-----NDVSLKAAR 1493
 QY 317 TIYADSRNVDEGRSTQSVLMSALDGFNIIRYQVFTFMNDKGEAIDKQGNLVTDSKLV 376
 DB 1494 TSASQAGSV---TKQTVLQG-----SANDKPYLLTLTGSKASSVKFT- 1534
 QY 377 FKQDKKEY-----TGDDKFNVEAIKEDGSMFLFTDKPVLNLSMDK 415
 DB 1535 ---DEKDYSLQANVTYADGSTGIYNAKFPSPQEWNRRAV-----VIPKTFKPNKVDIS 1585
 QY 416 NYFNPSKSNKIYVRNPEFYLRGKISDKGGFNELRVNESVDNLYLGLHDINTDRDFNI 475
 DB 1586 ILFQKSATGVW-----FD-----DIRLEGSLLTKSTYDSNG 1618
 QY 476 KLVNKGDIIMWGMK--DYKANGFPDPKVTMDG-----NVYLQTSYDLNA- 519
 DB 1619 NYVTKEEDELGYATSDYDTGKTKSETDAKGEKTTYTYDQADQLTNWLTNGTSILHSY 1678
 QY 520 -----KAGGVHVOFLYDNV--KPEVNIIDPKGNT--STIYADGKSVVFNINDKRNN 565
 DB 1679 DKEGNEVSKTRAGADQTYKFYDVMGKLVKTTDPLGNVLASEYDANSNLTKTISPNGNE 1738

QY 566 ---GPGETQEQHIYINGKEYTSFNIDIKQLIDKTLNLIKIVVDFAFNTTVKEFILNKOTG 622
Db 1739 VLSYDGTDRVRSKSYNGTEKIYFT-----YDKNGNETSVNKK-EQNTTKKRTFDNK--- 1789
QY 623 EVSELKPHRVTVTIQNGKEMSSIVSEED-----FILPVYKG-----E 660
Db 1790 -----NRLTELDRGGQTWTYPDSXKLTFSWIHGDKQKTNQFTYKNKLDOMIEMKD 1842
QY 661 LEKGYQFDGWE-----ISGFEKGDAGY-VINLSKDTFI-----KPVFKIEEKEE 707
Db 1843 STSYSDYDENGWQTFITNGGGTFSFYDERNLVSSLHIGDKNGDILTESYEYDANG 1902
QY 708 NKPTFDVSKKQNPQVNHQSLNESHKEDLOREHSQ-----KSD 747
Db 1903 NRTTIN-SSASGKVQVEYKGLNQ-----LVKETHEDGTVEIYDGFGRNRKTVTTIKDG 1955
QY 748 STKDVATV-----LDKNISSKSTNNPNK 773
Db 1956 SSKTVNASFNMQLVKVNDESISYDKNGNR 1986
RESULT 15
CNA_STAAU STANDARD; PRT; 1183 AA.
AC Q53654;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Collagen adhesin precursor.
GN CNA.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FDA 574;
RX MEDLINE=92165839; PubMed=1311320;
RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
RA Lindberg M., Hoeoek M.;
RT "Molecular characterization and expression of a gene encoding a
RT Staphylococcus aureus collagen adhesin."
RL J. Biol. Chem. 267:4766-4772(1992).
RN [2]
RP ERRATUM.
RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
RA Lindberg M., Hoeoek M.;
RL J. Biol. Chem. 269:11672-11672(1994).
RN [3]
RP COLLAGEN-BINDING DOMAIN.
RC STRAIN=FDA 574;
RX MEDLINE=94032261; PubMed=8218209;
RA Patti J.M., Boles J.O., Hoeoek M.;
RT "Identification and biochemical characterization of the ligand
RT binding domain of the collagen adhesin from Staphylococcus aureus."
RL Biochemistry 32:11428-11435(1993).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.
RX MEDLINE=97475225; PubMed=9334749;
RA Symersky J., Patti J.M., Carson M., House-Pompeo K., Teale M.,
RA Moore D., Jin L., Schneider A., Delucas L.J., Hoeoek M.,
RA Narayana S.V.L.;
RT "Structure of the collagen-binding domain from a Staphylococcus
RT aureus adhesin."
RL Nat. Struct. Biol. 4:833-838(1997).
CC -!- FUNCTION: MEDIATES ATTACHMENT OF STAPHYLOCOCCAL CELLS TO
CC COLLAGEN-CONTAINING SUBSTRATA.
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M81736; AAA20874.1; --
DR PDB; 1AMX; 24-JUN-98.
DR PDB; 1DZO; 27-SEP-00.
DR PDB; 1DZP; 27-SEP-00.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR006192; LPXTG.
DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; FALSE NEG.
KW Cell wall; Peptidoglycan-anchor; Repeat; Signal; 3D-structure.
FT SIGNAL 1 29
FT CHAIN 30 1154
FT PROPEP 1155 1183
FT DOMAIN 151 318
FT DOMAIN 533 1093
FT DOMAIN 1093 1157
FT REPEAT 533 719
FT REPEAT 720 906
FT REPEAT 907 1093
FT SITE 1151 1155
FT MOD_RES 1154 1154
FT STRAND 174 179
FT TURN 182 183
FT TURN 185 186
FT STRAND 187 194
FT TURN 196 197
FT STRAND 201 201
FT STRAND 205 211
FT STRAND 215 228
FT TURN 229 230
FT STRAND 232 234
FT HELIX 239 246
FT TURN 248 249
FT STRAND 251 255
FT TURN 256 259
FT STRAND 260 265
FT HELIX 267 270
FT TURN 271 272
FT STRAND 273 283
FT TURN 286 287
FT STRAND 290 299
FT STRAND 301 301
FT TURN 302 303
FT STRAND 307 311
FT STRAND 314 317
SQ SEQUENCE 1183 AA; 133066 MW; B6A1CC072E575D76 CRC64;
Query Match 4.5%; Score 179.5; DB 1; Length 1183;
Best Local Similarity 18.7%; Pred. No. 0.77;
Matches 172; Conservative 128; Mismatches 325; Indels 297; Gaps 42;
QY 22 KDDTTGVHHHQQENBESIKSSFTIDRNISTIRDFENKDLK-----KLKKKREV 73
Db 304 KEENVGKSFNHTVHNINAMAGIEGTVKGLVKL--QDKTKAPIANVVKLSKKDGSVV 361
QY 74 DDFSETGKRMEEYDYKDDKG--NIIAYDDGTDLEYETE-----KLDEIKSIYGVLS 126
Db 362 KD-----NQKEIITDANGIANIKALPSGDYILKKEIAPRPYTFDKKEYPFTMKDT 414
QY 127 SKDGHFEILGKISNVSKN-----AKVYGNKYKIEIKATKYDFHSTKTFTDLYANINDI 181
Db 415 DNOGFTFTIENAKAIEKTKDVSQAQWEGTQ-----KVKFTIY----- 452
QY 182 VDGLAFAGDMRLFLVKNDDQK-----KAEIKRMPEKIKETKSEYPVSSYGNVIEL---- 232
Db 453 -----FKLYKDDNQNTTTPVDKAEIKLEDDGTTKVTWNSLPENDKNGKAIKYLKVE 503
QY 233 --GGEDSKNKPNDNLTKMESKIIYSDSEKQVLLKNDNIIIRKGYALKVTTYNGKTDML 290
Db 504 VNAOGE--DTTPEGYTKKENGVLVNTTEK-----PIETTSIS 538

Search completed: January 28, 2004, 13:00:44
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.
OM protein - protein search, using sw model
Run on: January 28, 2004, 12:57:43 ; Search time 20 Seconds
(without alignments)
3716.920 Million cell updates/sec
Title: US-10-067-385-8
Perfect score: 4026
Sequence: 1 KLGETAESKFNKLGNGKEGS.....ATVLKNNISSKTTNNPNK 773
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : PIR 76: *
1: piri: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4026	100.0	2140	2 F95074	serine proteinase,
2	3709.5	92.1	2144	2 A97942	metalloproteinase,
3	253.5	6.3	1127	2 T28317	ORF MSV156 hypothe
4	250	6.2	1850	2 T18444	hypothetical prote
5	235.5	5.8	3724	2 T18427	hypothetical prote
6	231.5	5.8	2269	2 T28677	rhoptery protein -
7	230	5.7	2401	2 T28676	rhoptery protein -
8	228	5.7	4550	2 T18440	hypothetical prote
9	221.5	5.5	1711	2 T18429	hypothetical prote
10	219	5.4	1840	2 A24594	probable major sur
11	218	5.4	1631	1 SAZQK1	major merozoite su
12	218	5.4	1639	2 S05603	major merozoite su
13	217.5	5.4	1546	2 G90603	lipoprotein (impor
14	215.5	5.4	1252	2 B42771	reticulocyte-bindi
15	214	5.3	2339	2 A45597	DNA-directed RNA p
16	212.5	5.3	1125	2 E90598	membrane nuclease,
17	212.5	5.3	2166	2 G70163	hypothetical prote
18	212.5	5.3	2485	1 H71621	serine/threonine-s
19	212	5.3	1188	2 A71621	protein with 5'-3'
20	211.5	5.3	1104	1 A36866	microbial collagen
21	209	5.2	1558	2 F71603	RESA-H3 antigen PF
22	207.5	5.2	4688	2 E89824	hypothetical prote
23	207	5.1	1141	2 E89824	hypothetical prote
24	207	5.1	1191	2 B97116	chromosome segrega
25	207	5.1	1979	2 C71622	hypothetical prote
26	206.5	5.1	1622	2 AE1717	probable cell surf
27	206	5.1	1526	2 A45605	mature-parasite-in
28	204.5	5.1	1516	2 E71619	RAD2 endonuclease
29	204.5	5.1	2500	2 G71609	hypothetical prote

RESULT 1
F95074
serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: F95074
R:Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hei-
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfle,
non, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: F95074
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-2140 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK74791.1; PID:g14972117; GSPDB:GN00164; TIGR:SP
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP0841

ALIGNMENTS

Query Match	100.0%	Score 4026	DB 2	Length 2140
Best Local Similarity	100.0%	Pred. No. 4e-167		
Matches 773	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	KLGETAESKFNKLGNGKEGSLKDDTTGVHHHQHNEESIKESSTIDRNI	TIRDFFNK	60
DB	1334	KLGETAESKFNKLGNGKEGSLKDDTTGVHHHQHNEESIKESSTIDRNI	TIRDFFNK	1393
QY	61	DLKLLIKKKFREVDFTSETGKRMEEYDYKDDKGNIIAYDDGTDLVETEKLD	ETKSKI	120
DB	1394	DLKLLIKKKFREVDFTSETGKRMEEYDYKDDKGNIIAYDDGTDLVETEKLD	ETKSKI	1453
QY	121	YGVLSPSKDGFEILGKISNVSNAKYYGNNYKSIIEIKATKYDFHSTMT	FDLYANIND	180
DB	1454	YGVLSPSKDGFEILGKISNVSNAKYYGNNYKSIIEIKATKYDFHSTMT	FDLYANIND	1513
QY	181	IVDGLAFAGDMRLFKVNDQKAEIKIRMPKIKETKSEYPVSSYGNVIEIG	EGDLSKN	240
DB	1514	IVDGLAFAGDMRLFKVNDQKAEIKIRMPKIKETKSEYPVSSYGNVIEIG	EGDLSKN	1573
QY	241	KPDNLTKMESGKIYSDSKQQVLLKNDIILRGYALKVTTYPNGKTMLE	GNVYKEDI	300
DB	1574	KPDNLTKMESGKIYSDSKQQVLLKNDIILRGYALKVTTYPNGKTMLE	GNVYKEDI	1633
QY	301	AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNII	RYQVFTFMNDKGEA	360
DB	1634	AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNII	RYQVFTFMNDKGEA	1693
QY	361	IDKGNLVTDSSKLVFKDKKEYTGDKFNVEAIKEDGSMLEFIDTKPV	NLSMDKNYFNP	420

Db 1694 IDKGNLVTDSSKLVFGKDDKEYTGEDKFNVEAIKEDGSMFLFIDTKPVNLSMDKNYFNP 1753
Qy 421 SKSNKIYVRNPFYLRGKISDRGGFNWELRVNESVDVNYLIYGDHLIDNTRDNFKLNK 480
Db 1754 SKSNKIYVRNPFYLRGKISDRGGFNWELRVNESVDVNYLIYGDHLIDNTRDNFKLNK 1813
Qy 481 DGDIMDWGMKDYKANGFPDKVTDMDGNVYVLTQGYSDLNKAVGVHYQFLYDYNKPEVND 540
Db 1814 DGDIMDWGMKDYKANGFPDKVTDMDGNVYVLTQGYSDLNKAVGVHYQFLYDYNKPEVND 1873
Qy 541 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIQEHIYINGKEYTSFNDIKQIIDKTLN 600
Db 1874 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIQEHIYINGKEYTSFNDIKQIIDKTLN 1933
Qy 601 IVVKOPARNTTVKEFLNKDTEVSELKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYK 660
Db 1934 IVVKOPARNTTVKEFLNKDTEVSELKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYK 1993
Qy 661 LEKGQFDGWEISGPEGKDGAGVNLNLSKDTFKPVFKKIEEKKBEENKPTFDVSKKDN 720
Db 1994 LEKGQFDGWEISGPEGKDGAGVNLNLSKDTFKPVFKKIEEKKBEENKPTFDVSKKDN 2053
Qy 721 POWNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 773
Db 2054 POWNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 2106

RESULT 2
A:97942
C:metalloproteinase (RC 3.4.21.-) A [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: A97942
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: A97942
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2144 <KUT>
A:Cross-references: GB:AE007317; PIDN:AAK99365.1; PID:g15458138; GSPDB:GN00174
A:Gene: prtA
C:Genetics:
C:Keywords: hydrolase; serine proteinase

Query Match 92.1%; Score 3709.5; DB 2; Length 2144;
Best Local Similarity 92.4%; Pred. No. 2.2e-153;
Matches 714; Conservative 26; Mismatches 32; Indels 1; Gaps 1;
Qy 1 KLGEIAESKFNGLNGKESGLKDDTTGVHEHHQENESIKESSTFIDRNISTIRDFENK 60
Db 1339 KLGEISDGLKNAKSEFNT-NNQAKDESDKPEKSSVEGEASLEINKTISTIRFENK 1397
Qy 61 DLKLLIKKKFREVDDFTSETGRMEYDYKDDKGNIIAYDDGTDLEYETEKLDEIKSKI 120
Db 1398 DLKLLIKKKFREVDDFTSETGRMEYDYKDDKGNIIAYDDGTDLEYETEKLDEIKSKI 1457
Qy 121 YGVLSPSKDGHPFELIKISNVSKNAKYVGNVYKSTIEIKATKYDFHKTMTFDLYANIND 180
Db 1458 YGVLSPSKDGHPFELIKISNVSKNAKYVGNVYKSTIEIKATKYDFHKTMTFDLYANIND 1517
Qy 181 IVDGLAFAGDMRLFVNDQKAEIKIRMPKIKETKSEVPVYSSYGNVIELCEGDLN 240
Db 1518 IVDGLAFAGDMRLFVNDQKAEIKIRMPKIKETKSEVPVYSSYGNVIELCEGDLN 1577
Qy 241 KPNLTKMESGKIYSDSEKQYLLKDNIIILRKGVALKVITYNPKGTDMLGNGVSKEDI 300
Db 1578 KPNLTKMESGKIYSDSEKQYLLKDNIIILRKGVALKVITYNPKGTDMLGNGVSKEDI 1637

Qy 301 AKIQANPNLRALSETTIYADSRNVEDGSRSTQSVLMSALDGFNIIIRYQVFTFMNDKGEA 360
Db 1638 AKIQANPNLRALSETTIYADSRNVEDGSRSTQSVLMSALDGFNIIIRYQVFTFMNDKGEA 1697
Qy 361 IDKGNLVTDSSKLVFGKDDKEYTGEDKFNVEAIKEDGSMFLFIDTKPVNLSMDKNYFNP 420
Db 1698 IDKGNLVTDSSKLVFGKDDKEYTGEDKFNVEAIKEDGSMFLFIDTKPVNLSMDKNYFNP 1757
Qy 421 SKSNKIYVRNPFYLRGKISDRGGFNWELRVNESVDVNYLIYGDHLIDNTRDNFKLNK 480
Db 1758 SKSNKIYVRNPFYLRGKISDRGGFNWELRVNESVDVNYLIYGDHLIDNTRDNFKLNK 1817
Qy 481 DGDIMDWGMKDYKANGFPDKVTDMDGNVYVLTQGYSDLNKAVGVHYQFLYDYNKPEVND 540
Db 1818 DGDIMDWGMKDYKANGFPDKVTDMDGNVYVLTQGYSDLNKAVGVHYQFLYDYNKPEVND 1877
Qy 541 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIQEHIYINGKEYTSFNDIKQIIDKTLN 600
Db 1878 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIQEHIYINGKEYTSFNDIKQIIDKTLN 1937
Qy 601 IVVKOPARNTTVKEFLNKDTEVSELKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYK 660
Db 1938 IVVKOPARNTTVKEFLNKDTEVSELKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYK 1997
Qy 661 LEKGQFDGWEISGPEGKDGAGVNLNLSKDTFKPVFKKIEEKKBEENKPTFDVSKKDN 720
Db 1998 LEKGQFDGWEISGPEGKDGAGVNLNLSKDTFKPVFKKIEEKKBEENKPTFDVSKKDN 2057
Qy 721 POWNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 773
Db 2058 POWNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 2110

RESULT 3
T28317
ORF MSV156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus
C:Species: Melanoplus sanguinipes entomopoxvirus
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T28317
R:Afonso, C.D.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.
J. Virol. 73, 533-552, 1999
A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.
A:Reference number: Z20484; MUID:99102612; PMID:9847359
A:Accession: T28317
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1127 <AFO>
A:Cross-references: EMBL:AF063866; NID:g4049647; PIDN:AAK97677.1; PID:g4049717
C:Genetics:
A:Note: MSV156

Query Match 6.3%; Score 253.5; DB 2; Length 1127;
Best Local Similarity 21.5%; Pred. No. 0.0011;
Matches 186; Conservative 155; Mismatches 338; Indels 185; Gaps 44;
Qy 7 ESKEKNLNGKESGLKDDTTGVHEHHQENESIK--EKSSFTID-RNISTIRDFENKDLK 63
Db 196 ELEFNKNDVQKEINNKQDELNKLDESKEFKKQELNKTIDKQELIKKLNDEIN 255
Qy 64 KLIIKKKFRVDDFTSETGRME-----EYDYKYDDKGNIIAYDDGT--DLEYETE 111
Db 256 FNIDEKQLDDQINSKINTLNENIKGVNMLYTETKNKLNQELNKLNDSTIKSLDEKQK 315
Qy 112 KLDEIKSKIYGVLSPSKDGHPFELIKISNVSK-----NAKYVGNVYKSTIEIKAT 161
Db 316 LLELDKNNINNTSILYNKSN---TKITNIOQLLESSLTFNNANT---NINELSKIK 367
Qy 162 KYDFHKTMTFDLYANINDIYDGLAFAGDMRLFVK--DNDOKKA-EIKIRMPKIKET-- 216
Db 368 LFDNDIQKUNNDITEONNKITD--FPNNSTRIFKEKLDTEYKKIDDKNNLOKLESYK 425
Qy 217 ----KSEY---PYVSSYGNVIELCEGDLSKNKPDLNLTVMESGKIYSDSEKQYLLKDNII 269

Db 426 KIDEQTEYKYNKINKEYNIDIELKNNLQKLEENKKIDEQTEYKYNKINKEY--NDIIE 483
Qy 270 LRKGYALKVTTNPG-----KTMLEGGVGYKEDIAKQKPNLRALSETTIYAD 321
Db 484 LKNNLQKLEENKINIDKLTAKNDIESNTELFNKLINISDFPKOKSREIAKLN--TEYEQ 541
Qy 322 SRN--VEDGRSTQSVL-----MSALDGNRIYQVFTFMNDKGAIDKGNLVTDSK 373
Db 542 LRKDLLENINKTNELMKLSDNKLSSLE-----QLYDSKKNIL-DGIDKIYNSLKEKN- 592
Qy 374 LVLFGKDKKEYTCEDKFNVEAIKED--GSMFLFIDTPVNLSDMKNYFN--PSKNKIY 427
Db 593 -----DKIDEVFSNIEKFDIYVNIENKFIKINLSDSIINKIINNOFKIYINSKIDSNSLS 648
Qy 428 VRNPEFY-----LRKISDKGGFNWELRVNYSVDNLYIYGLIHIDNTRD 472
Db 649 TWFDDIFNAKNOIASITNNIENISNKID-----LNEFIISNEDSSKEL-LDEIRK 698
Qy 473 FNIKLVKQGDIMDGMKDYKANGFPDKVTDMDGNV-YLOTGYSDLNAKA-----V 522
Db 699 YKQOQD-KIDAMNTEVKSFE-NTLQKIDISIKSNINELTNAYDIINTKANLDDKLNNY 756
Qy 523 GVHYQFLYDNVXP-----EWNIDPKGNTSIEVAD-CKSVVFNINDKRNNGFDEIQEQHI 576
Db 757 GSEFKNLNINASDLDTIOKNNDKVKQLNEYLEKONQOSIEINDIVNN-----805
Qy 577 YINGKEYTSFNIDKIQIDITLNIKIVVDFARNTTVKFEILNKDTGEVSE---LKPHRYT 633
Db 806 FI--KELIKFNTE--TNKSLN-ELLTND--DINDKIFKLYKELNKISTNNLLKIY--855
Qy 634 VTIONKEMSSITVSEEDFILPVYKGELEKGYQFDGWEISGEGKADGVINLSKDTF- 692
Db 856 NEIDNVNEXLSIVLENLOFINSFLSTEFNOG-----SITSH-----INFLNTLA 900
Qy 693 -IKPVFKTLEEKKEENKTFEDVSKKNDPQVNHSQLNESHKEDLQREHSQKSDSTKD 751
Db 901 GINDVLNKLNLKIMADTTRRGDTNIR-----DEIKNOISSENKSKOFNEKNB--KD 950
Qy 752 VTATV-----LDKNNISSKSTNN 770
Db 951 LKKLISFNDKLNKYNISAGYTEYN 974

RESULT 4

T18444
hypothetical protein C0385c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T18444
R:Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z18935
A:Accession: T18444
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1650 <LAW>
A:Cross-references: EMBL:Z98547; NID:e1325376; PID:e1427940; PIDN:CAB11112.2
C:Genetics:
A:Map position: 3
A:Introns: 1597/3; 1625/3
A:Note: C0385c

Query Match 6.2%; Score 250; DB 2; Length 1650;
Best Local Similarity 20.8%; Pred. No. 0.0026;
Matches 205; Conservative 134; Mismatches 320; Indels 328; Gaps 48;
Qy 20 SLKQDTGVHHEHQEENEEISIKSSFTIDRNFENKOLKLIKFKFRE---VDDF 76
Db 394 SLNENYNERKIYKEINKEEYSNKEVY-----HFKND-DSSIKKKNSSSECLDRQ 443
Qy 77 TSETGQR--MEEVDYKYDDKGNLIAY--DDGTDL-----YETKLDIKSKIYGVISPSKD 129
Db 444 KKKTYKTYIEQKRYNFNDNRNNNAVYIKDTHKKEGYLLNMTVQSEYKTYG--SNRKM 501

Qy 130 GHFEILGKIS-----NVSKNAKYVYGNNYKSI--EIKATKYDFHSKTMTFDLYANINDIV 182
Db 502 DEMEINYQHTNDFNINENLNKIYF--DDYEGVDPEKKKKLDDHIYTOQKEYKNNINDIL 560
Qy 183 DGLAFAGDMRLFKVNDOKKAEIKIRMPKIK-----ETKSEYPYVSSVGNVIELEG-- 235
Db 561 K-----DHLNDKETKEKKNIEIEEKKKNKIEIEEKKKNKIEIEEKKKNKIEIEEKK 614
Qy 236 ---DLSKNKPDNLTKMESGKIYSDSEKQOYLKDNII-----LR 271
Db 615 KKIEIEEKKKKI-EMEEBKNDIDDEKONTYANDKIISHIDNVNCNIKIDALLDHIEKK 673
Qy 272 KGYALKVTTYNPGKTD-----MLEGGVYSKEDIAK-----IOKANPUL-- 310
Db 674 KTGHKEINLYKEIKNEYQKMLNDENSIMLEHEKKNYTHQVNNNLCDTKMLQKENKILT 733
Qy 311 RALSETTIYADSRNVEDGRSTQSVLMSALDGNRIYQVFTFMNDKGAIDKGNLVT 370
Db 734 NDKKKTFLLSKSKNI-----TSNVLSSKIPG-----TLSTKCLNATIITIKD--VTD 779
Qy 371 SSKLVLFQ-----KODKE---YTGEDKE-NVEA-----IKEDGSMFLFIDTK 407
Db 780 NEKKYVVDHRKDNIIKRNKEFINIYKGNANVANVEIGSEVCNKNVNVKDDNQMVENK 839
Qy 408 P-----VNLSDMKNYF-----NPSKSNKIYVRNPEF 433
Db 840 QGDDNNMVMVENKQDNNVIVKNIERSSESSFMFTHRKNITSTSTDCAKNKEIQKIYPHL 899
Qy 434 YLRGKISDKGGFNWELR-VNESVDNY-----LYDGLHIN-----469
Db 900 Y-SNRKEDKOKKSIIFLKNINENIKKNYKOKEKISTLEKKVFVKQDNVITNDDERTSSK 958
Qy 470 -TRDFNI-----KLN-----478
Db 959 INDDFNITIDKQKGLNPNVLDNRIKIKETKLEKDKSHMSKIQNNLMKKTNFTNNK 1018
Qy 479 -VKDGDIMDGMKDYKANGFPDKVTDMDGNVYL-----QTG 513
Db 1019 GISSTSISSSTKFNKFCGIIIEKNKLSNLKYTCIRKNKNVDSIKLNDKADLYDKKKTS 1078
Qy 514 YSDLNKAVGHVHYQFLYDNVNVKPEVNIDPKGNTSIEYADOKSVVFN-----INDKRNNGFDG 569
Db 1079 FNDINRAAKGMNFK-----KRDV---PNKNNMVDNFKGR-VFNPVTLNNNYRNN---1124
Qy 570 EIQEQHIYINGKEYTSFNIDKIQIDITLNIKIVVDFARNTTVKFEILNKDTGEVSELKP 629
Db 1125 -----YIRSNK-----NNVK-----NGRW-----GIKKILVLKE--KQKSLHP 1156
Qy 630 HRVTVTIONGKEMSSITVSEEDFILPVYKGELEKGYQFDGWEISGEGKADGVINLSK 699
Db 1157 EGVEADKKLNSYNDKYLIEKDGFKDIIINEEKE-----YKNNK-MKYKI--K 1201
Qy 690 DTFIKPVFKKIEEKEENKPTFDVSKKNDPQVNHSQLNESHKEDLQREE---HSQKS 746
Db 1202 SNSIPIIKKIERKSNNDN---DNKKNNDNINSNNNINSNDKCLFLSKERDVRHLKNN 1258
Qy 747 DSTKDVTATVLDKNNISSKSTNNPNK 773
Db 1259 NIIVANTMMFRKQSNSCDNNTTSLKNN 1285

RESULT 5

T18427
hypothetical protein C0335c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T18427
R:Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z18935
A:Accession: T18427
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-3724 <LAW>
A:Cross-references: EMBL:Z98547; NID:e1325376; PID:e1325379; PIDN:CAB11104.1
C:Genetics:
A:Introns: 307/1; 1545/2
A:Note: C0335c

Query Match 5.8%; Score 235.5; DB 2; Length 3724;
Best Local Similarity 18.7%; Pred. No. 0.032;
Matches 192; Conservative 150; Mismatches 312; Indels 371; Gaps 45;

QY 5 IAESKFNLCNGKGS-LKK-----DTTGVHEHGOEESIK-----BKSSFT 46
DB 249 IPKNLENLNNKHNDYLRNILLVDNDINPLEHPDEQNSLNKCLTGNKKKYM 308
QY 47 I-----DRNSTIRDFENKDKLIIKKKPREVDFTSETGKRMEEDYKYDDKG 95
DB 309 IPKGTBYQDKEKESILTINQNDKKYKKK-KKKYSELQDSNISSNNTLTSKRYTYT--- 363
QY 96 NIIAYDDGTDLEVEKLEDEIKSIYGVVSPSKDGHFEILGKISNVSKNAKYVGNKYK 155
DB 364 -----CGMDKETKETDEQNK-----RNTEVASICNVCVNETNEKORKKTKKK 407
QY 156 IEIKATKYDFHSKTMFTFDLYANINDIVGLAFAGDMR--LFVKDNDQKKAIEIKIRMP-EK 212
DB 408 ---SVQKDLVEENALLD---NL-IIIDGINFDDVTKECKIIDNNNNENIDNNIYENNK 460
QY 213 IKETKSEYPVSSYG-NVI-----ELGEGDLSKNKPDNLTKMESGK 252
DB 461 LKDKQS-YDLFSSEGNLILGVNEGEEFNEFENIEKELQERKDEKKNK-----T 512
QY 253 IYSDSEKQVLLKDNILIRKGVALKVTTYNPGKTMLEGNVYSGEDIAKIQAN-PNLR 311
DB 513 IYNNNEEOTDLNRNI-----NKIESINNNNDN---NNNINNKKEFNKIRTEHLNKE 562
QY 312 ALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVTFKMDKGEAIDKDGMLVTS 371
DB 563 SISKHKIGSPSRDK-----IKKLYTNKNESTFELKKELEIITNN 604
QY 372 SKLVFGKDKYTGEDKFNVEAIKEDGSMFLFDITKPVNLSMDKNVFNPSKN----- 424
DB 605 KVNIVY-EBDIIIGSNEDDEVIHLKE-----NLKEDANEYNDKENKKNKKE 650
QY 425 ----KIYVRN-----PEFYLRGKISDKGFNWLNVN---ESVDNYLYIGDLHIDTDD 472
DB 651 ILKSKNYLENKRTLEELKRGK---NNIFKDEKYNLSGEVLIINEIQINEENKINDIQ 707
QY 473 ENIK-----LNVKD---GDIMDMGMDKYKANGFPDKVTDMD----- 505
DB 708 GNISKOKIIQSSRTWDTFNKIDISLNDLLEKREKKKSKQHFDNLVAKDKNEISENINK 767
QY 506 -----GNVYLQT-----GVSOLNAXAVGVHY----- 526
DB 768 ICDNNINNIYDESINNIYDESINNIYDESINNIYDENINNIYDENINNIYDB 827
QY 527 -----QFLYDNVPEVNID-----PKNTSIE-----Y 549
DB 828 NINNIYDEGINKICDDNILENKNKIKTNDIVQVEENNESIEKNELMISLNKOINNYYNMF 887
QY 550 ADGKSVFN-----INDKRNNGFDGEIOE----- 573
DB 888 KENVDFINKIKRESLLKDKDNKONNDDEYIMDNFYNDFINHMKBITNKELDPLEI 947
QY 574 -----OHYINGKEYTS---FNDIKQIIDKTLNIIKVVKDFARNVTYKFEILNKDTGB 623
DB 948 NTQNEFIENLDIKKKYTYNDHPFNADAKMPYE-MN-KILNKKMKKKEQEFKTDFTFGS 1005
QY 624 VSELK-----PHRVTVTIQNGKMSSTIVSE 649
DB 1006 LQSHKIKKYNKGEKHKDNNEEKNIYDENOVSVLYSDHKTEODIQIHSIQTNICDE 1065
QY 650 EDFILPVYGELEKGYQFDGWEISGPEGKADAGVINLSKDTFKPVFKKIEKKEENK 709

DB 1066 NN-IEQINEENS KGVRIISGDTM-----ENKND-----MENKNDMEKK 1102
QY 710 PTFDYVSKKNDPOVNH-SQLNESHRKEDLQREHSQKSDSTKDVATVTLVDKNISKSTT 768
DB 1103 N--DMEKKNDEKKNDEKKNDEKKNDEKKNDEKKNDEKKNDEKKNDEKKNDEKKNDE 1153
QY 769 NNPKN 773
DB 1154 ENENK 1158

RESULT 6
T28677
rhoptry protein - Plasmodium yoelii
C:Species: Plasmodium yoelii
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
A:Accession: T28677; C45521
R:Keen, J.; Sinha, K.; Brown, K.; Holder, A.
Mol. Biochem. Parasitol. 65, 171-177, 1994
A:Title: A gene coding for a high molecular mass rhoptry protein of Plasmodium yoelii.
A:Reference number: Z20508; MUID:95021522; PMID:7935623
A:Accession: T28677
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2269 <KE2>
A:Cross-references: EMBL:IL27838; NID:9457145; PID:9457146; PIDN:AAA21304.1
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A:Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple c
A:Reference number: A45521; MUID:91101660; PMID:2270106
A:Accession: C45521
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 2131-2269 <KE2>
A:Cross-references: GB:M34283

Query Match 5.8%; Score 231.5; DB 2; Length 2269;
Best Local Similarity 19.7%; Pred. No. 0.025;
Matches 181; Conservative 147; Mismatches 330; Indels 259; Gaps 37;

QY 37 ESIEKKSFTIDRNTSTIRDFENK-DLKKLIKKKPREVDFTSETGKRMEY----- 87
DB 109 EGLKLELNKKIIIAKIEYKNTVELKEIEKNAYIDELANOSPKYKTYIENKNTTY 168
QY 88 -----DYKYDDKGNIIAYDDGTDLEVEKLEDEIKSIYGVVSPSKDGHF 132
DB 169 NTIKSYFOIYEGDITDFYNELSSIVKEDPIDIDEDKT-KLENLRKSDINVDKIQKMEI 227
QY 133 E-ILGKISNVSKNAKYVGNKYKSIKATKYDFHSKTMFTFDLYANINDIVGLAFAGDM 191
DB 228 ETVVSHLNNIETNNKL-----PNTILEIKYIYDEISK-----ELNKWLEDFKNK---EKEL 276
QY 192 RLUFKNDQDK--AEIKIRMP-----KIKETKSEYPVVSSVGNVJELGE 234
DB 277 SNKISDYDKRQLESEYKSKMLEIRNHNVSQTNVDNTEKEEAQKQYDKSNEHMTTPTNE 336
QY 235 GDLK-----NKPDNLTKMES-----GKIYSD-----SEKQVYL-LKONI---ILRKYAL 276
DB 337 DEISKIIISVTKMDKBEILSKVNTYIDFNKYYKETVNSEHSQFTELTDKIKAEVSDKELKK 396
QY 277 KVTYNPGKTMLEGNVYSGE--DIAKIQANPNLRALSET-----TIYAD-- 321
DB 397 CQSFNDNKSILNETKNSIEKYQNITLKKVDEYIKVCKSTKESITKFSKQTLKMDL 456
QY 322 SRNVEDGRSTQSVLMSALDGFNIIRYQVET-----FKNN-----DKGEAI--- 361
DB 457 NQNIKTVKETNSIDKSYIEKFE---QILTGRQTKLENKFTFESLANNEANNELIKYFS 512
QY 362 KDKGHLVTDSSKLVLFGKDDKEYTGEDKENVEAIKEDGSMFLFDITKPVNLSMDKNVFNPS 421
DB 513 DLKANLGINNEENLWYNQFTEKETFND-----IKENIHIINEEISKIEIKHASIYNIS 566
QY 422 KSKNIYVRNPEFYLRGKISDKGFNWLNVSNVDNYLYIGDLHIDTRDFNKLNVK- 480


```
QY 175 YANINDIVDGLAPAGDMRLFVNDQKKAIEIKRMPEKIKETKSEYPPYVSSYGNVIELGE 234
Db 268 Y--IEKIDOLFMAST--IKDTYO-----TNKIFLTNDY----- 298
QY 235 GDLKSNKPNLTKME-----SGKYS-----DSEKQOYLLKDNILIRKGYAL 276
Db 299 --IKNNESNLKDKWFSNNFENIYKYFLEKLDALVEIQALHKQYLEKNEIILKTEIVD 356
QY 277 KVTYTPNGKTDLENG--YYSKEDIKI-----QKANPNLRALS 314
Db 357 KVAAPFKSRELKSKAGLLFSKQDSAKITQLINHILSRYPPEAPENLELSKAKWGLINEL 416
QY 315 ETIIVADSRNVEDGRST--QSVLMSALDGNFIIRYQVTFPMKMDKGEAIDKGNLVTDSK 373
Db 417 KPEIYVEKEGVIAYPTLODAISNAQDGQKIFL-----NKNLKLK--SIVYDKN- 464
QY 374 LVLFGKDKKEYTGED--KFNVEALIEDGSMFLFDTKPVNLSMDKNYFNS----- 421
Db 465 ITIFAKSNVTIIRKOSKSFTHFIQKGAITFEIAPESQSNLNGLGTSFKDESSLVKI 524
QY 422 -KSNKIYVRNPFYLRGKISDRGGFNWELRVNESVVDNYLYIGDLHIDNTRDN----- 474
Db 525 EKNAKLVAKTGTAFINSKFSKYG-----SVFEN--YGSVIEGAKIWNVSESG 572
QY 475 -----IKLVKGDIMDMGMDKYKANGFPDK-VTDMGDNVYLOTGYSDNLNAKAVGVH 525
Db 573 GIIRNVGSSSLTFKNGEI-----RDNISGT--DKGIYSGQNIATSGSIDGNKS----- 620
QY 526 YQFLYDNVKEPVNIDPKNGTSEYADGKSVFNINDKR-----NNGFDGELQSHIYING- 580
Db 621 FRSSSLINLE-KTNINFNSGSIVNNAASVKISILFEIDNSKIOISNNALINPFGSSAIFLKN 679
QY 581 -----KEYTSFNIDIKQIIDKTLNLIKIV----- 602
Db 680 STMHLAGSLKIKKEASEQRIEVLVLPQAKLISPKNIISLDNYQLSSAIFKIFSVKN 739
QY 603 VKDPAR-----NTTVKEFILNKDT-----GEVSELKPHRVTVTIQNGKEMSTIVSEB-DF 652
Db 740 INDFKHVPLVMNTKEKFFKLWPDTKLVFNFYKTLKQNDLILQSGDGFESTEKIIKDELDF 799
QY 653 -----ILPVYKLEKGYOF-DGWEISGEGKK-----DAGYVINLSK 689
Db 800 YFRPTAAVKKILITQLVRTIPKTHKFWAEFVNPDLQKWETINELIRLDPFYLFDIAY 859
QY 690 DTFIKPVPFKIBKEKEEENKPTFDSKKDNPOVNSQLNSHRKEDLQR 739
Db 860 PEFV-----ENGKMLKPEY-IHTNVNVPVLEH-----FRNEDVAR 894

RESULT 14
B42771
reticulocyte-binding protein 2 - Plasmodium vivax (fragment)
C:Species: Plasmodium vivax
C:Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 09-Sep-1997
C:Accession: B42771
R:Galinski, M.R.; Medina, C.C.; Ingravallo, P.; Barnwell, J.W.
Cell 69, 1213-1226, 1992
A:Title: A reticulocyte-binding protein complex of Plasmodium vivax merozoites.
A:Reference number: A42771; MUID:92315338; PMID:1617731
A:Accession: B42771
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1252 <GAL>
A:Cross-references: GB:M88098; NID:gl60627; PID:gl60628
A:Experimental source: strain Belem, merozoites
C:Genetics:
A:Gene: RBP
```

Query Match 5.4%; Score 215.5; DB 2; Length 1252;
Best Local Similarity 20.9%; Pred. No. 0.056;
Matches 207; Conservative 130; Mismatches 324; Indels 331; Gaps 50;
QY 2 LGIEAESKPK--NLNGKSGSLKDKDTTGVGHHQNEESIKESST-IDRNI----- 51

RESULT 15

A45597

DNA-directed RNA polymerase (EC 2.7.7.6) III largest chain - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jun-2000

```
Db 266 LSEIKYKDKCTTEISNKR-----KDKTEFLKFPNEESNKNVINEINIRNSEQY 322
QY 52 -----STIRDFENKDLKLIK--PREVDDFTS 78
Db 323 LKDIEDAQAQSTKVLPFKHETTTISNIFKSESEILGVETKSQKINKAEDIMKEIERHNS 382
QY 79 ETGKMEEYDYKDDKGNIIAYDDGTDLEYEYETKLEDEIKSKIYGVLSPSKDGHEFLGKI 138
Db 383 ETIQTKVGQFQENLNKLNPHNYDNAED--ELNNDKSTNAKVLJETNLESVKHN-----LSEI 437
QY 139 SNVSKNAKVYGNVYKSI--EIKATKYDPHSHKTM-----TFDLYAN-INDIVDGLAFAGD 190
Db 438 TNIKOGGEKIY--SKAKDIMQKIKATSENTAEKTELEKVKDDQSNVYNYLQIIT-----E 490
QY 191 MRLFKVND-----OKKAETKIRMEPIKE-----TK-----S 218
Db 491 RNLIVTEKRNLRNGIDSTITNIEGALKESGNYEIGFLEKLEIGKRNKLVKVIDTKKSINS 550
QY 219 EYPYVSSYGNVIELGEGDLSKKNPNLTKMESKGIYSDEKQOYLLKDNILIRKGYALKV 278
Db 551 TVGNFSSLPNFDLQYDFNKNINDYENKM--GEIYNEPEGSINKISENL--RNASENT 605
QY 279 TTYNPGKTDMLEGNGVYSKEDI---AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVL 335
Db 606 SDYNSAKTLRLA-----QKEKYNLLNKEEANKYLR-----DVKKVESFR----- 646
QY 336 MSALDGFNIIRYQVTFKMDKGEAIDKGNLVTDSKLVLFKGDQKKEYTGEDKFNVEAI 395
Db 647 -----FIFNMK-----ESLDKINEMIKKEQLTV-----NEGHGNVKQLVENI 683
QY 396 KEDGSMFLFDTKPVNLS-----MDKN-----YFNPSKNKIYVRNPFYLRGKI 439
Db 684 KE-----LVDEN--NLSDLKQATGKNEEIQKITHSTLKNKAKTILGHVDTSAKYVGIKI 736
QY 440 SDKGFNVELRVNESVVD-----NLYIYCDLHIDNTRDNFKLNKVD--- 481
Db 737 TP-----ELALTTELLGDAKLTQAELKFPESKNVYLETENMSKNTNELDVHKNIQDAYK 790
QY 482 -----GDIMDMGMDKYKANGFPDKVTDMDGNVYL-----QTGYSDLNAKAVGVH 525
Db 791 VALEILAHSDIEDTKQD-----SSKLIEMGNQIYLVKVLINQYKNKISSKSEEAVS 844
QY 526 YQFLYDNV-----PEVNIDPKNGTSEYADGKSVFN-----NDKNNNGFDGELQIOBQ 574
Db 845 VKI--GNVSKHSELKITSCKSDSYDNIILAEKQTELQNLRSFTQEKNTNDSKLE-- 900
QY 575 HIYINGKEYTSFNDIKQIIDKTLNLIK-----VVKD----- 605
Db 901 -----KIKTDFESLKNAL-KTLEGEVNAKASSDNHEHVQSKSPVNPALSEIEKEETD 953
QY 606 -PARNTTVKEPFLNKDTGEVSELKPHRVTVT--IQNGKEMSTIVSEEDFILPVYKGELE 662
Db 954 IDSLNTALDELLKKGRTCEVSRYKLIKOTVTKEISDDTELINTIEKN-----VKAYLAYIK 1009
QY 663 KYQFDGWEISGEGKDGAGYVINLSKDTFIKPV-----FKKIEEKEEENKPTFD-- 713
Db 1010 KNYE-----DTVQD--VLTLEHEFNTKQVSNHEPTNFDKNGKSSEELTKAVTDSK 1057
QY 714 --VSKKK-----DNQPVN-----HSQLNESHR-KEDLQREHESQ 744
Db 1058 TTISKLGVIIEVNTENTMTTIESAKIEALYNELKNKNTSLNEIYOTSNEVKLQEMKS 1117
QY 745 KSDSKVDVTA--TVLD--KXNISKSTTN 770
Db 1118 NADKIDVSKIPFNTVLDTKQSKNIVNQHSINN 1149
```

C:Accession: A45597
R.Li, W.B.; Bzik, D.J.; Tanaka, M.; Gu, H.M.; Fox, B.A.; Inselburg, J.
Mol. Biochem. Parasitol. 46, 229-239, 1991
A:Title: Characterization of the gene encoding the largest subunit of Plasmodium falciparum
A:Reference number: A45597; MUID:92018020; PMID:1656254
A:Accession: A45597
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2339 <L11>
A:Note: sequence extracted from NCBI backbone. (NCBIN:61099, NCBIPI:61101)
C:Superfamily: Plasmodium DNA-directed RNA polymerase II large chain
C:Keywords: nucleotidyltransferase; transcription

Query Match 5.3%; Score 214; DB 2; Length 2339;
Best Local Similarity 21.6%; Pred. No. 0.15;
Matches 191; Conservative 134; Mismatches 306; Indels 254; Gaps 50;

QY 24 DTTGVHHHQQE--NEESIKESKSTIDRNISTIRDP---ENKDLKKLIKKEPREVDDFTS 78
DB 1292 DNTYVEQIEKELSKNKTKEKQSPK-----GTIRDMHEDSEBQMNKFTTKAK----FFI 1342
QY 79 FTGKRMEEYDYKDKGNIIAYDDGTDLEYTEKLEDEIKSIY-----G 122
DB 1343 EK-KGKGMEHCNDIEYNTQYD---NIQYNNISCNVYKSNLENTHQVNNDLSPFKNN 1398
QY 123 VLSPSKDGH-----FELGKISNVSKNAKYVYNNYKSIKATKYDFHKTMTFPL 174
DB 1399 VILPPKEYHSIFHFVNDYRVVVEIKLMDKKKIFLNNSEKN--VVOVKYRMRMSKNLKKKI 1456
QY 175 YANINDIVDG-----LAPAGDMRLFVKNDQ--KKAETKIRMPK-----IKETSE 219
DB 1457 EI-INNIYRNEKKLNKRWTKMDNDNYWSSDDDSIIAKIIKIKNKEKKYHPKEKEN 1515
QY 220 YPVVSYGNVIELGEGDLKSNKPDNLTK-MESGKIYSDSEKQOYLKDNILIRKGYALKV 278
DB 1516 FDR-NNYKMITDNNNDNNNNNDNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1569
QY 279 TTYNPKTDML--EGNGVYSKEDIKIQKANPL-----RALSETTIYADS-- 322
DB 1570 TNYN---TNIYPNDCNGIYEKETNNNELTNSNMCNKNDFDEFFNNINENDLLYDNKY 1626
QY 323 -----RNVEDGRSTQVLSALDGFNIIRYQVF---TFKMDKGEAIDKGNLVTDSKL 374
DB 1627 YRQIFKNVIGFVSFVYVESYKQHYILFPYEIIKWTSPLELYTEIPTNIFLHTKLSK- 1685
QY 375 VLFGKD---DKYTGDEKFNVEAIEKDGSMFLFDTKPVNLSMDKNYFNPSKKNKIYVRNP 431
DB 1686 ---KEKPTHQKNTGKMKIYIEIKK---WLF--KAINI---YKVFSPKKSIEL----- 1728
QY 432 EFVLRGKISDKGFGNHELARNESVDNLYIGDLHIDNTRD-----FNI-----KLVKD 481
DB 1729 -----IKKQYFNYIIK-NYDISHRYIIH-DYSPINLKQLYLFYFFNIYKYFYISTP 1779
QY 482 GOIMDMGMDYKANGFPDKVTDMDGNVYLTQGYSDLNKAVGVHYQFLYDNVKNPEVNIDP 541
DB 1780 GD----AVGSIASAQSGEPTQTLKTFHFAGVASMNV-TLGV-----PRIKEIINASN 1828
QY 542 KGNITSIEYAGKSVFVNINDKRNNGF-----DGEIQOQHVIYNGKEYTS-----F 586
DB 1829 SIQTPI-----LNIPLEVNDNYNFALMMSKLEKTTIRDCMYIK-EDYTSRGVFLSVKF 1882
QY 587 ND-----IKQIIDKTLNI-----KI---VVKOPARNTVK--EFILNKDT 621
DB 1883 NEELIQKLFININAYNIKDIILKQSHINKIKINKIHINVINKYKHLHISLKNDEFIF---- 1938
QY 622 GEVSELKPHRVTVTIQNGKEMSTIVSEDFILPVYKGELEKGYQDFGWGEISGFEKG--K 679
DB 1939 FQMESLKGLGLDLLIYGDKDIKECIKKEDI-----EVTNDEDEICD 1980
QY 680 DAGYVINLSKDTFKPVFKIEKK---BEENKPTFDVSKKDNPOVNHSQLNESHKED 736
DB 1981 DMDEYTNVSGT-----ELYERKCNSEENKNAIRVKKEID-----DN 2019

QY 737 LQREE-----HSQKSDSTKDVATATVLDKNNISSKSTNN 770
DB 2020 LEKEENIIYVSEKDSVQNLKSEKKDIN-----DDNNNDNNNNN 2060
Search completed: January 28, 2004, 13:02:11
Job time : 25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 12:56:43 ; Search time 42 Seconds
(without alignments)
4749.397 Million cell updates/sec

Title: US-10-067-385-8
Perfect score: 4026
Sequence: 1 KUGETAESKFNKNGKEGS.....ATVLKNNISSKTTNNPNK 773

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4026	100.0	2119	2 Q9AHT5	Q9AHT5 streptococc
2	4026	100.0	2140	16 Q37RY6	Q37RY6 streptococc
3	3793.5	94.2	2144	2 Q9S4M8	Q9S4M8 streptococc
4	3709.5	92.1	2144	16 Q8DQP7	Q8DQP7 streptococc
5	265.5	6.6	1642	5 Q8IB84	Q8IB84 plasmodium
6	264	6.6	3504	5 Q8IL45	Q8IL45 plasmodium
7	255	6.3	1777	5 Q8I3P4	Q8I3P4 plasmodium
8	253.5	6.3	1127	12 Q3YVT6	Q3YVT6 melanoplus
9	252.5	6.3	2227	5 Q8II21	Q8II21 plasmodium
10	251.5	6.2	2849	5 Q8IH4	Q8IH4 plasmodium
11	251.5	6.2	3381	5 Q8I2V4	Q8I2V4 plasmodium
12	251.5	6.2	3519	5 Q8IE65	Q8IE65 plasmodium
13	250	6.2	1474	5 Q8IIU2	Q8IIU2 plasmodium
14	250	6.2	1650	5 Q7J328	Q7J328 plasmodium
15	247.5	6.1	5767	5 Q8I525	Q8I525 plasmodium
16	246.5	6.1	2033	5 Q8IM18	Q8IM18 plasmodium

17	245	6.1	2269	5 Q8ILA2	Q8ILA2 plasmodium
18	242.5	6.0	3317	16 Q8EWP8	Q8EWP8 mycoplasma
19	241.5	6.0	1033	5 Q8IBB8	Q8IBB8 plasmodium
20	240.5	6.0	4433	5 Q8IJ15	Q8IJ15 plasmodium
21	240	6.0	1455	5 Q8IKG8	Q8IKG8 plasmodium
22	240	6.0	10061	5 Q8I3Z1	Q8I3Z1 plasmodium
23	239.5	5.9	2569	5 Q8IBG8	Q8IBG8 plasmodium
24	238	5.9	1081	16 Q8XIL2	Q8XIL2 clostridium
25	238	5.9	3322	5 Q8IKL0	Q8IKL0 plasmodium
26	238	5.9	3628	5 Q968Y0	Q968Y0 plasmodium
27	238	5.9	3704	5 Q8IKY8	Q8IKY8 plasmodium
28	237.5	5.9	3535	5 Q8IC29	Q8IC29 plasmodium
29	237	5.9	2511	5 Q8IL44	Q8IL44 plasmodium
30	236.5	5.9	1114	5 Q97242	Q97242 plasmodium
31	236.5	5.9	1738	5 Q8IAL5	Q8IAL5 plasmodium
32	236.5	5.9	4044	5 Q8IDD4	Q8IDD4 plasmodium
33	236	5.9	2228	5 Q8I1W5	Q8I1W5 plasmodium
34	236	5.9	3620	5 Q968T6	Q968T6 plasmodium
35	235.5	5.8	1692	5 Q8IJD6	Q8IJD6 plasmodium
36	235.5	5.8	3724	5 Q77320	Q77320 plasmodium
37	235	5.8	1936	5 Q8ISA6	Q8ISA6 plasmodium
38	235	5.8	2184	5 Q8ILO6	Q8ILO6 plasmodium
39	234	5.8	2082	5 Q8I2R1	Q8I2R1 plasmodium
40	234	5.8	2359	5 Q8I297	Q8I297 plasmodium
41	234	5.8	2760	5 Q8ISY2	Q8ISY2 plasmodium
42	233.5	5.8	2752	5 Q9BJY0	Q9BJY0 plasmodium
43	233.5	5.8	3610	5 Q968T7	Q968T7 plasmodium
44	232.5	5.8	7170	5 Q8IL30	Q8IL30 plasmodium
45	232	5.8	2361	5 Q8IEJ4	Q8IEJ4 plasmodium

ALIGNMENTS

RESULT 1

Q9AHT5 PRELIMINARY; PRT; 2119 AA.
ID Q9AHT5;
AC Q9AHT5;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Serine protease (fragment).
GN PRTA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N4;
RX MEDLINE=21116976; PubMed=11179332;
RA Witzmann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,
RA Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,
RA Gayle A., Brewah Y.A., Walsh W., Warren P., Lathigra R., Hanson M.,
RA Langermann S., Johnson S., Koenig S.;
RT "Use of a Whole Genome Approach to Identify Vaccine Molecules
RT Affording Protection against Streptococcus pneumoniae Infection.";
RL Infect Immun 69:1593-1598(2001).
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY
CC AN AMIDE BOND (BY SIMILARITY).
EMBL AF291699; AAK19159.1; -.
HSSP: P00782; 2SPT.
DR MEROPS: S08.064; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR006192; LPXTG.
DR InterPro: IPR003137; PA.
DR InterPro: IPR002029; Peptidase_S8.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR Pfam: PF02225; PA; 1.
DR Pfam: PF00082; Peptidase_S8; 2.
DR PRINTS: PR00723; SUBTILISIN.
DR TIGRFAMs: TIGR01167; LPXTG_anchor; 1.

DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS50840; PA; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
KW Cell wall; Peptidoglycan-anchor; Protease.
FT NON_TER 1
SQ SEQUENCE 2119 AA; 238226 MW; 517F9B756B960A6A CRC64;

Query Match 100.0%; Score 4026; DB 2; Length 2119;
Best Local Similarity 100.0%; Pred. No. 6.6e-154;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLGEIAESKFNGLNGKESLKKDTTGVHHHQENEESEIKKSSFTIDRNISTIRDFENK 60
DB 1313 KLGEIAESKFNGLNGKESLKKDTTGVHHHQENEESEIKKSSFTIDRNISTIRDFENK 1372

QY 61 DLKLLIKKKFREVDFTSETGKMEEDYKDDKGNIIAYDDGTDLLEYETEKLEIKSKI 120
DB 1373 DLKLLIKKKFREVDFTSETGKMEEDYKDDKGNIIAYDDGTDLLEYETEKLEIKSKI 1432

QY 121 YGVLSKDGHEIFLKGISNVSKNAKYVYGNVYKSIIEIKATKYDFHSHKTMFDLYANIND 180
DB 1433 YGVLSKDGHEIFLKGISNVSKNAKYVYGNVYKSIIEIKATKYDFHSHKTMFDLYANIND 1492

QY 181 IVDGLAFAGDMRLFVKNDQKAEIKIRMPKIKETKSEYPYVSSYGVNIELEGDSKN 240
DB 1493 IVDGLAFAGDMRLFVKNDQKAEIKIRMPKIKETKSEYPYVSSYGVNIELEGDSKN 1552

QY 241 KPDNLTKMESGKIYSDSEKQVLLKDNIIILRGYALKVTTYNPCKTMDLEGNGVYSKEDI 300
DB 1553 KPDNLTKMESGKIYSDSEKQVLLKDNIIILRGYALKVTTYNPCKTMDLEGNGVYSKEDI 1612

QY 301 AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFMNDKGEA 360
DB 1613 AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFMNDKGEA 1672

QY 361 IDKGNLVTDSSKLVFGKDDKEYTGEDKFNVEAIKEDGSMFLFIDTKPVNLSMDKNYFNP 420
DB 1673 IDKGNLVTDSSKLVFGKDDKEYTGEDKFNVEAIKEDGSMFLFIDTKPVNLSMDKNYFNP 1732

QY 421 SKSNKIYVNPPEYLRGKISDKGFFNWLRVNESVVDNLIYGLDHDNTRDFNKLNVK 480
DB 1733 SKSNKIYVNPPEYLRGKISDKGFFNWLRVNESVVDNLIYGLDHDNTRDFNKLNVK 1792

QY 481 DGDIMDWKMDYKANGFPDKVTDMDGNVYLQGYSDLNKAKAVGVHVFQFLYDNVKNPEVND 540
DB 1793 DGDIMDWKMDYKANGFPDKVTDMDGNVYLQGYSDLNKAKAVGVHVFQFLYDNVKNPEVND 1852

QY 541 PKGNTSIEYADGKSVVFNINDKNNNGFDGEIQEOHIYINGKEYTSNDIKQIIDKTLNFK 600
DB 1853 PKGNTSIEYADGKSVVFNINDKNNNGFDGEIQEOHIYINGKEYTSNDIKQIIDKTLNFK 1912

QY 601 IVVKDFARNTVKEFILNKDTGVBSELKPHRVTVTTIQNGKMSSTIVSEDFILPVYKGE 660
DB 1913 IVVKDFARNTVKEFILNKDTGVBSELKPHRVTVTTIQNGKMSSTIVSEDFILPVYKGE 1972

QY 661 LEKGYQFDGWEISGFGKDGAGVYINLSKDTFIFKPVFKIIEKKEBENKPTFDVSKKXON 720
DB 1973 LEKGYQFDGWEISGFGKDGAGVYINLSKDTFIFKPVFKIIEKKEBENKPTFDVSKKXON 2032

QY 721 PQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 773
DB 2033 PQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 2085

RESULT 2
Q97RY6 PRELIMINARY; PRT; 2140 AA.
AC Q97RY6;
DT 01-OCT-2001 (T-EMBLrel. 18, Created)
DT 01-OCT-2001 (T-EMBLrel. 18, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)

DE Serine protease, subtilase family.
GN SP0641
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TIG84;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
R Durkin A.S., Gwin M., Kolonay J.F., Nelson W.C., Peterson J.D.,
R Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
R Holtzapple E., Kouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
R McDonald L.A., Feldblyum T.V., Anguoli S., Dickinson T., Hickey E.K.,
R Holt I.E., Lotius B.J., Yang F., Smith H.O., Venter J.C.,
R Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae";
RL Science 293:498-506 (2001).
DR EMBL; AE007373; AAK74791.1; -.
DR MEROPS; S08.064; -.
DR TIGR; SP0641; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR006192; LPXTG.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 2.
DR PRINTS; PR00723; SUBTILASIN.
DR TIGRFAM; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS50840; PA; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
KW Protease; Complete proteome.
SQ SEQUENCE 2140 AA; 240426 MW; FA44AD8E2938B334 CRC64;

Query Match 100.0%; Score 4026; DB 16; Length 2140;
Best Local Similarity 100.0%; Pred. No. 6.6e-154;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLGEIAESKFNGLNGKESLKKDTTGVHHHQENEESEIKKSSFTIDRNISTIRDFENK 60
DB 1334 KLGEIAESKFNGLNGKESLKKDTTGVHHHQENEESEIKKSSFTIDRNISTIRDFENK 1393

QY 61 DLKLLIKKKFREVDFTSETGKMEEDYKDDKGNIIAYDDGTDLLEYETEKLEIKSKI 120
DB 1394 DLKLLIKKKFREVDFTSETGKMEEDYKDDKGNIIAYDDGTDLLEYETEKLEIKSKI 1453

QY 121 YGVLSKDGHEIFLKGISNVSKNAKYVYGNVYKSIIEIKATKYDFHSHKTMFDLYANIND 180
DB 1454 YGVLSKDGHEIFLKGISNVSKNAKYVYGNVYKSIIEIKATKYDFHSHKTMFDLYANIND 1513

QY 181 IVDGLAFAGDMRLFVKNDQKAEIKIRMPKIKETKSEYPYVSSYGVNIELEGDSKN 240
DB 1514 IVDGLAFAGDMRLFVKNDQKAEIKIRMPKIKETKSEYPYVSSYGVNIELEGDSKN 1573

QY 241 KPDNLTKMESGKIYSDSEKQVLLKDNIIILRGYALKVTTYNPCKTMDLEGNGVYSKEDI 300
DB 1574 KPDNLTKMESGKIYSDSEKQVLLKDNIIILRGYALKVTTYNPCKTMDLEGNGVYSKEDI 1633

QY 301 AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFMNDKGEA 360
DB 1634 AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFMNDKGEA 1693

QY 361 IDKGNLVTDSSKLVFGKDDKEYTGEDKFNVEAIKEDGSMFLFIDTKPVNLSMDKNYFNP 420
DB 1694 IDKGNLVTDSSKLVFGKDDKEYTGEDKFNVEAIKEDGSMFLFIDTKPVNLSMDKNYFNP 1753

QY 421 SKSNKIYVRNPEYLRGKISDKGGFNWELRVNESVVDNYLYGDLHIDNTRDFNKLNVK 480
DB 1754 SKSNKIYVRNPEYLRGKISDKGGFNWELRVNESVVDNYLYGDLHIDNTRDFNKLNVK 1813
QY 481 DGDIMDGMKDYKANGFPDVKVTDMDGNVYLTQGYSDLNKAKAVGVHYQFYLDNVKPEVND 540
DB 1814 DGDIMDGMKDYKANGFPDVKVTDMDGNVYLTQGYSDLNKAKAVGVHYQFYLDNVKPEVND 1873
QY 541 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIQEHIYINGKEYTSFNDIKQIIDKTLNLIK 600
DB 1874 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIQEHIYINGKEYTSFNDIKQIIDKTLNLIK 1933
QY 601 IVVKDFARNTTVKEFILNKDTGEVSELKPHRVTVTTIQNGKEMSSSTIVSEDFILPVYKGE 660
DB 1934 IVVKDFARNTTVKEFILNKDTGEVSELKPHRVTVTTIQNGKEMSSSTIVSEDFILPVYKGE 1993
QY 661 LEKGYQFDGWEISGFGKDGAGVNLKSDTFLKPVFKKIEBKKEENKPTFDVSKKDN 720
DB 1994 LEKGYQFDGWEISGFGKDGAGVNLKSDTFLKPVFKKIEBKKEENKPTFDVSKKDN 2053
QY 721 POWNHSQLNESHKEDLOREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 773
DB 2054 POWNHSQLNESHKEDLOREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 2106

RESULT 3

Q9S4M8 PRELIMINARY; PRT; 2144 AA.
AC Q9S4M8;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Cell wall-associated serine proteinase PrtA precursor.
GN PRtA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN (1)
RC STRAIN=3.B;
RA Bethel G., ten Thoren E., Bongaerts R.J.M., Heinz H.-P., Zysk G.;
RT "Cloning and sequencing of a novel surface protease of Streptococcus pneumoniae."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY AN AMIDE BOND (BY SIMILARITY).
CC EMBL; AF127143; AAD48399.1; -.
DR HSP; P00782; 2SBT.
DR MEROPS; S08.064; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR006192; LPXTG.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 2.
DR PRINTS; PR00723; SUBTILISIN.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00840; PA; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
KW Cell wall; Peptidoglycan-anchor; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 2144 CELL WALL-ASSOCIATED SERINE PROTEINASE PRtA.
SQ SEQUENCE 2144 AA; 240724 MW; 2052511470741331 CRC64;

Query Match

94.2%; Score 3793.5; DB 2; Length 2144;

Best Local Similarity 94.7%; Pred. No. 1.5e-144;
Matches 732; Conservative 16; Mismatches 24; Indels 1; Gaps 1;
QY 1 KLGEIAESKFKNLGNKGKSGIKKDTTGVHHHHEEESIKESKSFITDRNISTIRDFENK 60
DB 1339 KLGEIPESKFKNLGNKDDSLNKETAENVNLLVDNQSLGSLFNHKTISTIRDFENK 1398
QY 61 DLKULIKKKPREVDFTSETGRMBEYDYKYDDKGNIIAYDDGTDLLEYTEKLEIKSKI 120
DB 1399 DLKULIKKKYQEDDFVN-GGTRTVVERDYKYDDKGNIIAYDDGTDLLEYTEKLEIKSKI 1457
QY 121 YGVLTSPSKDGHPEIILGKISNWSNKAQVYGNKYKIEIKATKYDFHSTKMTFDLYANIND 180
DB 1458 YGVLTSPSKDGHPEIILGKISNWSNKAQVYGNKYKIEIKATKYDFHSTKMTFDLYANIND 1517
QY 181 IVDGLAFAGDMELFVKDNDQKAEIKIRMPKIKETKSEYPYVSSYGNVIELEGDSLN 240
DB 1518 IVDGLAFAGDMELFVKDNDQKAEIKIRMPKIKETKSEYPYVSSYGNVIELEGDSLN 1577
QY 241 KPDNLTKMESGKIYSDSEKQVLLKDNILIRKGYALKVTYTPGKTDMLBEGNVYSKEDI 300
DB 1578 KPDNLTKMESGKIYSDSEKQVLLKDNILIRKGYALKVTYTPGKTDMLBEGNVYSKEDI 1637
QY 301 AKIQANPNLRALSETTIYADSRNVEDGSRSTOSVLMALDGENIIRYQVTFKMDKGEA 360
DB 1638 AKIQANPNLRALSETTIYADSRNVEDGSRSTOSVLMALDGENIIRYQVTFKMDKGEA 1697
QY 361 IDKGNLVTDSKLVLFKDDKEYTGEDKFNVEALIKEDGSMFLFIDTKPVNLSMDKNYFNP 420
DB 1698 IDKGNLVTDSKLVLFKDDKEYTGEDKFNVEALIKEDGSMFLFIDTKPVNLSMDKNYFNP 1757
QY 421 SKSNKIYVRNPEYLRGKISDKGGFNWELRVNESVVDNYLYGDLHIDNTRDFNKLNVK 480
DB 1758 SKSNKIYVRNPEYLRGKISDKGGFNWELRVNESVVDNYLYGDLHIDNTRDFNKLNVK 1817
QY 481 DGDIMDGMKDYKANGFPDVKVTDMDGNVYLTQGYSDLNKAKAVGVHYQFYLDNVKPEVND 540
DB 1818 DGDIMDGMKDYKANGFPDVKVTDMDGNVYLTQGYSDLNKAKAVGVHYQFYLDNVKPEVND 1877
QY 541 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIQEHIYINGKEYTSFNDIKQIIDKTLNLIK 600
DB 1878 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIQEHIYINGKEYTSFNDIKQIIDKTLNLIK 1937
QY 601 IVVKDFARNTTVKEFILNKDTGEVSELKPHRVTVTTIQNGKEMSSSTIVSEDFILPVYKGE 660
DB 1938 IVVKDFARNTTVKEFILNKDTGEVSELKPHRVTVTTIQNGKEMSSSTIVSEDFILPVYKGE 1997
QY 661 LEKGYQFDGWEISGFGKDGAGVNLKSDTFLKPVFKKIEBKKEENKPTFDVSKKDN 720
DB 1998 LEKGYQFDGWEISGFGKDGAGVNLKSDTFLKPVFKKIEBKKEENKPTFDVSKKDN 2057
QY 721 POWNHSQLNESHKEDLOREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 773
DB 2058 POWNHSQLNESHKEDLOREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 2110

RESULT 4

Q8DQ7 PRELIMINARY; PRT; 2144 AA.
AC Q8DQ7;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Cell wall-associated serine proteinase PrtA (EC 3.4.21.-).
GN PRtA OR SPR0561.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,


```
Db 692 DLNKRKSNIDIQVNNNNNNNDTTFYDAMEYSGNDKEQYKNVQBD-HVINGVESKNKN 750
QY 681 -----AGYVNLKSDTFI-----KPVF-- 697
Db 751 MOTNCSSNNKYNINDRWNHLKYEYDMKPGVDNIIINIKENELVLQKNEKKDIFNK 810
QY 698 -----KKIEKEEENK--PTFDVSKKDDNPQVNH-----SOLNESH-----RKEDLQRE 740
Db 811 SHGNEKIEPKDFNNKRVKVTNLKKNVNDIHEDEVKSVNIGNEIDENRKIKKENILNS 870
QY 741 EHSKSDSKD-VTATVLDKNNISSKSTNNPN 772
Db 871 INQEKDIGKNIIKNTSYSKNLLSKYAEKPH 903

RESULT 6
Q81L45
ID Q81L45 PRELIMINARY; PRT; 3504 AA.
AC Q81L45;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Hypothetical protein.
GN Pf14_0404.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22555705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Fertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511 (2002).
DR EMBL; AE014822; AAN37017.1; -.
KW Hypothetical protein.
SQ SEQUENCE 3504 AA; 408303 MW; B8454D48D55BE4F0 CRC64;

Query Match 6.8%; Score 264; DB 5; Length 3504;
Best Local Similarity 21.2%; Pred. No. 0.018;
Matches 196; Conservative 131; Mismatches 346; Indels 252; Gaps 43;

QY 11 KNLGNCKGSLKDTTGVVHHQEN--EESIKEK-----SFTIDRNISTIRDPENKDL 62
Db 1432 KNSHENSENIMTEYKGNKNSPOENIDNNIPEKNGENKNSQNSDHNIMTEKNGKNS 1491
QY 63 KKLKLLKFFREVDFTSETGRMBEYDYKYDDK----GNIIAYDDGDTLEYETEKLDEIKS 118
Db 1492 QQ-----NSDHNIMTEYKGNKNSQENTDNNIMTEYKGNKNSKNEKEEDIAS 1539
QY 119 KIYGV-----LSPSKD-GHEILGKINSVKNKAYVG-----NNYKSIKATKYDF--- 165
Db 1540 --YEIDKNRISHENDQEHFTPESRINKGFHTNVYVNSGNEEKGISVKNISEDIIPD 1597
QY 166 -HSKWTFTFLYANINDVGLAPAGDMFLPVKNDQKAEIKIRMEKIKETKSE---YP 221
Db 1598 GKGKNIQDDIIILEN-----GENKPFENIEEDKISDKTQ-KSKISHENAEGHFTP 1647
QY 222 YVS-----SYGVNIELGEGDLNKPNDLTKMESGRIY-----SDSEKQOYLL 264
Db 1648 YESGKNKISDENDVEYNSDINTNKDQE--EVESKEIFETNDINKNGHSSDNNKINKMK 1705
QY 265 KONILRKGVALKVYTNYPGKTD-MLEGNGVSKEDIKIQKANPNLR-----ALSETTI 318
```

```
Db 1706 QNNILINESQDKNIQDVHN--KLDKILKNEHVTSDSLEKIKEENGTRKNKGSINNEEKI 1763
QY 319 YADSRNV-----EDGRSTQSVLMSALDGFNIIRYQVFTFKNN----- 355
Db 1764 BEEKENVKNDETIIGKKEENESDDLKIQKISNENINKNILYDNTYKDKSYNAQGGTHG 1823
QY 356 -----DKGEAID-----KGNLV----- 368
Db 1824 ENDETTNGTNIISGDLKDNVKIDQISGENILOENKEDIIPSVTINNSLGDYKVKENLS 1883
QY 369 -TDSKLVLFQKDDKEYTGEDKFNV-----EAKEDGSMFLFDTK-----PVNLS- 412
Db 1884 PEDIKKMEVAHKNIQNIITSEDELGTQKDNERNKEDKSPNGVEENHQENDKIIGEVLNS 1943
QY 413 MDKNYFNPSKSNKIYVRNPEFYLRGK-ISKDGKGFNWEI--VNESVVDNY-----L 460
Db 1944 MNMESNIGNSDTI---NOHLLNEGKNIHKKGNVNSSETNMTNNSGTQNIISNEQFEKNI 2000
QY 461 IYGDLIHIDNTRDFNKLAVKQGDIMDMGMYKANGFPDKVDTMDGNVYVLTQGYSLNAK 520
Db 2001 IRGD-DIKDKMNVNKIEDETGN-----NIKINKYNDNAKILNELIINKOGTQDSAD 2052
QY 521 AVGVHYQFLYDNVK-PEVNIDPKGNTSLEYADGKSVFVNINDKRNNGPDEIQEQHIYN 579
Db 2053 DISTNGSKMDQIENKNNENIHKNNVVE-----KDKI--SNDKRNIVPPE-HKEDHILS 2105
QY 580 GKEYTSFNDIKQIIDKTLNLIKIVKDFARNTTVKFELNKDTCGEVSELKPHRVTVTIQNG 639
Db 2106 DNKKKEFDNVLEIPKGNH-----LDDKETITEQVEKSIQODKSMENNVT--NDG 2157
QY 640 KMSSTIVSEEDFILPVYKGELEKGYQDFGWEISGFEGKKDAGYVINLSK----- 689
Db 2158 KDIIH---IQEEDI-----KGNIIINVNDKHSKKNLHIDEPNKYVEEKEIKKH 2204
QY 690 ----DTFIKPVFKKIEEKEENKPT-----FDVSKKKNPQVNSHLSHREKEDLQREE 741
Db 2205 ETADHDIKKEFKIEQENSNKNEPSNENILVDVNAQDDK-----NISKLTDNLHDOE 2256
QY 742 HSKQSDSKDVTATVLDKNNISKS 766
Db 2257 KGTNDSVVE-----HNVS DKT 2273

RESULT 7
Q813P4
ID Q813P4 PRELIMINARY; PRT; 1777 AA.
AC Q813P4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Hypothetical protein.
GN PFE1095W.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Devlin K., Baker S., Davies P., Mungal K., Berriman M., Pain A.,
RA Hall N., Bowman S., Churcher C., Quail M., Barrell B.;
RL Submitted (SEP-2002) to the EMBL/Genbank/DBSJ databases.
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22555708; PubMed=12368867;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C., Clark R.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kethornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,
```

RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrell B.G.
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).
DR EMBL; AL929353; CAD51583.1; --
KW Hypothetical protein.
SQ SEQUENCE 1777 AA; 213320 MW; 244467CFPF190C522 CRC64;

Query Match 6.3%; Score 255; DB 5; Length 1777;
Best Local Similarity 20.7%; Pred. No. 0.019;
Matches 186; Conservative 152; Mismatches 324; Indels 236; Gaps 44;

QY 4 ETAEKFKNLGKESLKKDTTGVHHQENE-----ESTKESSTIDRNISTIRDF 57
DB 977 ELLDDKKKL--DEENELDDKK--KKLDEENELDDKKKLEENELDDKKKL--DE 1030
QY 58 ENK--DLKKLI-----KKKPREVDFTSETGKMEYDYKDKGNLIAYDDGT 104
DB 1031 ENELDDKKKLEENELDDKKLDEENILLEERKKKMDENILLDEKKKEIVNDHNT 1090
QY 105 DLEYETEKLDEIKSKI---YGVLSPS-KDGHF-----EILGKISNVSKNAK-- 146
DB 1091 FIQTEHNLNKKTKLOEDYNIIEAELEKWHKSKILLLEETKIKIENDENDIKRKQSQIEN 1150
QY 147 VYGGNNYKSIIEIKATYDFHSTMTFDLYANINDIVDGLAFAGDMRLFKVNDQKKAETK 206
DB 1151 IYRRNSMWDININSYKSNVKTTFENFDNIEE----- 1184
QY 207 IRMPEKIKETKSEYPVSYGVNIEGEGDLKSKPDNLTKMESGKIYSDSEKQVLLKD 266
DB 1185 ---EDKITQKNTKYTDDISNM-----LTKNKKSSYISNEXKDIINEQEIYSNKN 1234
QY 267 NI-----ILRGYALKVTTNPGKTMLEGNVYSK-----EDIAKIQ-ANPN 309
DB 1235 NTFNNDIEKDLTNVMEKNFNINSYKQCSNIYDINDSNKFLMDTESIIKQINDINEK 1294
QY 310 LRALS--ETTYADSRNEDGRSTQSLM--SALDGFNIIRYQVTFKMDKGEAIDKGN 366
DB 1295 VKLLKQREEIFYBEKKNFKEKNEHLLENVLNKNWIIKDR--ENELNKEBELKMKKE 1352
QY 367 LVTDSSKLVLF-----GKDKKEY-----TGEDKFNVEAIKEDGSMFLFDITKPNLSM 413
DB 1353 ILLSKENIINYSLSNKKINKKLEYNKLEVDTKGGINMTSMKSFKSHL--SDTN--NIOP 1409
QY 414 DKNYFNPSKNIYVRNPFYLRGKISDKGGFNWELRV-----NESVDNLIY-G 463
DB 1410 FKNDIN--NDKLYEDNISSY--GNLSHDNNMYDNNLSCDNNMSHDNNWARDNNICHN 1465
QY 464 DLHIDNTRDFNIKLVN-----KDGIMDMGMDKYKANGPDPKVTMDGVY----- 509
DB 1466 NISYDNNPSYNTHMKLSELENNENKNKNIDY---YDKDNFVDRNNIINHNNFNISKI 1521
QY 510 ---LQGYSDLNKAVGVHYQFYDYNKPEVNDPKGNTSIEYADGKSVFVNIDKRNNG 566
DB 1522 SNDISGNMDMB-----LHDNL-----NNSLKEIEMYSKLSLSRSDSEINT 1562
QY 567 FDGEIOEQ-----HYINGK-----EY-TSFNDIKQIIDKTLN 598
DB 1563 LKLKIEKQTEKWKQNNINQNLNDISSDNNIYINGNISGVNYDTSMDVLELPERKFLN 1622
QY 599 IKIVKDPARNTVTKFILNKOTG---EVSLEKPHRVTVTIQNGKMSSTIVSEEDFI-- 653
DB 1623 LS-EVED---NEKWSILSNELSVLYKEINNIK-----EYNINVLKNEFIGN 1667
QY 654 -LPVYGELEKGYQFGWELSGEGKADAGVNLNSKDTFIKVPFKIEKKEEENKPTF 712
DB 1668 LLNLFNLDLRNNYRL---KENYYEKEVHKSQIIISEREDFIKELQNTLANEKKLKE----- 1719
QY 713 DVSKKDNPOVNSQLNESH---KEDLQREHSQKSDSTKDTATVLDKNNISKST 767

DB 1720 -ISYKMKLLKWN--QINDTYKLKNRSLSTVELLQK--DIKFLNEDVLKKEMVTLST 1772

RESULT 8
QYVTV6 PRELIMINARY; PRT; 1127 AA.
AC QYVTV6;
DT 01-MAY-1999 (TREMELrel. 10, Created)
DT 01-MAY-1999 (TREMELrel. 10, Last sequence update)
DE 01-OCT-2002 (TREMELrel. 22, Last annotation update)
DE ORF MSV156 hypothetical protein.
GN MSV156.
OS Melanoplus sanguinipes entomopoxvirus (MsEPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;
OC Entomopoxvirus B.
OX NCBI_TaxID=83191;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Tucson;
RX MEDLINE=99102612; PubMed=9847359;
RA Afonso C.L., Tulman E.R., Lu Z., Oma E., Kutish G.F., Rock D.L.;
RT "The genome of Melanoplus sanguinipes entomopoxvirus.";
RL J. Virol. 73:533-552(1999).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=Tucson;
RA Afonso C.L., Tulman E.R., Lu Z., Oma E., Kutish G.F., Rock D.L.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF063866; AAC97677.1; --
KW Hypothetical protein.
SQ SEQUENCE 1127 AA; 134265 MW; F185DA1D5A3FE7D1 CRC64;

Query Match 6.3%; Score 253.5; DB 12; Length 1127;
Best Local Similarity 21.5%; Pred. No. 0.013;
Matches 186; Conservative 155; Mismatches 338; Indels 185; Gaps 44;

QY 7 ESKFNLGNGKSGSLKDDTTGVHHQENEESIK--EKSFTID-RNISTIRDFENKOLK 63
DB 196 EIEFNIDNVQKEINKKQDELNKLDDSKKEIKKEELNKTIDKQBELIKKLNDKEIN 255
QY 64 KLIKPKPREVDFTSETGRME-----EYDYKDDKGNIIAYDDGT--DLEYETE 111
DB 256 FNIDEQKLLDQINSKINTLNENIKGVNMLYTETKNISNLQNEILNKDSTIKSLDEKQ 315
QY 112 KLEIKSKYIGVLSKDGHPILGKISNVSK-----NAKVVYGNVYSIKRIKAT 161
DB 316 LLDELDPKNNTISLYNKSNI---TKITNIQQLLESSLTFNNANI---NINELSKIK 367
QY 162 KYDFHSTMTFDLYANINDIVDGLAFAGDMRLFVK--DNDOKKA-EIKIRMPKIKET-- 216
DB 368 LFENDIQKLNNDITEQNKITD--FPNNSTRIFKEKLDTEYKKIDDIKNNLQKLEESYK 425
QY 217 ----KSEY---PYVSYGVNIEGEGDLKSKPDNLTKMESGKIYSDSEKQVLLKNDII 269
DB 426 KIDEQTEYVKNKINKEYNDIIEKNNNLOKLEENKKIDEQTEYVKNKINKEY--NDIIE 483
QY 270 LRKGVALKVTTPG-----KIDMLEGVYSKEDIKQKAPNLRALSETTIYAD 321
DB 484 LKNNLQKLEENKINDKLTKLNDIESNTELFKNLINSDFKDKSREIAKLN--TEYEQ 541
QY 322 SRN--VEDGRSTQSVL-----MSALDGFNIIRYQVTFKMDKGEAIDKGNLVDSSK 373
DB 542 LRADLLENKNTNMLKSLSDNKLSSLE-----QLYDSKKNIL-DGIDKIYNSLAKEN- 592
QY 374 LVLFGKDDKEYTGEDKFNVEAIKED---GSMFLFDITKPNLSMDKQVFN---PSKNNKIY 427
DB 593 ----DKIDYFVSNIEKFDIYVNIENKFIKNLDSIINKIINNDFQKEYINSKIDSKSNELS 648
QY 428 VRNPEFY-----LRKISDKGFNFWELRVNESVDNLYIGDLHIHNTD 472
DB 649 TWFDDIFNAKQIATISNTNIENISNKKID-----LNFEIISNEDSSKEL-LDEIRK 698
QY 473 FNIKLVNKGDIWDGMDKDYKANGPDPKVTMDGVN-YLQGTGYSDLNKA-----V 522


```
Db 699 YKQDF-KIKDAMNTEVKGFE-NLQKQDSIKSNINELTAYDIINTKANDLDDKLNNY 756
Qy 523 GVHYQFLYDNVXP-----EVNIDPKGNTSIEYAD-GKSVVFNIDKRNNGFQGEIQEHI 576
Db 757 GSEFKLYNNASDLLLOTIQKNDKVKQLNEYLEKKNQOSIENDIVNN-----805
Qy 577 YINGKYTFNDIKIIDKTLNLIKIVVQDFARNTTVKEFLNKGDTGEVSE---LKPFRVT 633
Db 806 FI--KELIKFNTE--TNKSLN-ELLTND---DINDKIFKLYKELNKKISTNNLLTIYK-- 855
Qy 634 VTIONKEMSSIVSEEDFLPVYKGELEKGYQDFGWEISGFGKKGAGVNLKDTF- 692
Db 856 NEIDNVNKLKSVIENLQFINSFLSIEFNQ-----SITSH-----INFLNTLA 900
Qy 693 -IKPVFKIEEKEEENKPTFDVSKKKNQPNVHNSQLNESHKEDLQREHSOKSDSTKD 751
Db 901 GINDVLNKLNLKIMATDTRGDINR-----DEIKQISSENIKSQKFNKNE--KD 950
Qy 752 VVATV-----LDKNISSKSTNN 770
Db 951 LKKLISFNDKLNKYNISAGYTEYN 974

RESULT 9
Q81121 ID Q81121 PRELIMINARY; PRT; 2227 AA.
AC Q81121;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF11_0354.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlson J.M., Pain A., Nelson K.E., Bowman S., Faulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Nature 419:498-511(2002).
DR EMBL; AE014841; AA35938.1; -.
KW Hypothetical protein.
SQ SEQUENCE 2227 AA; 262840 MW; 6D5D4BEEF643339C CRC64;

Query Match 6.3%; Score 252.5; DB 5; Length 2227;
Best Local Similarity 19.1%; Pred. No. 0.031;
Matches 190; Conservative 155; Mismatches 319; Indels 329; Gaps 49;

Qy 10 FKNLNGKE-----GSLKQDTTGVH-----HHQENESIEKEK 42
Db 690 YNNNNNNKCTTSINIKSHENKYPFNKSHVNSYMKNTNHLPHRNAITSNNRNEEYKEK 749
Qy 43 SSFTIDRNTST-----IRDPENK-----DLKLI-----KKKPREVDFT-----SE 79
Db 750 EK---DRNTGNNNNLYVEVNNSSCIPPLKMKIPIDGVNRKNSINKLNNTVNTQRTSSVS 806
Qy 80 TGRMEEDY-----KYDDKGNIIAVDGTDLEVEYETKLEIKSIYGVLSPSK 128
Db 807 TKNDIENSFDMPINGIESKIYSNNNI---NGNISGNSKLDNYHQSNV-----858
Qy 129 DGHFEILGISVSKNAKVYGNKYKSIIEIKATKYDFHSKWTTF-----DLYANIN---179
```

```
Db 859 -----NESYPLKMMKQNY-----IEHNYDDKNIFLVKQYEDTYSNIHNGI 900
Qy 180 -----DVGDLAFAGDME-----LFPVKON-----DOKK-----202
Db 901 HENSMKLYNLKACCTFHGYSRHHQKQMYTEENLNINQKQYSHYHNGTGLVKPLVNTNN 960
Qy 203 -----AETIKRMEKIKETKS-----EYPYVSSYGVNIELGEGDLSKQKPDNLTKMESGKI 253
Db 961 VAVNEFADINLSAQKRLHSLKSMGYEDKSMENYRNKIYNNINNNNNNDN-----NI 1013
Qy 254 YDSEKQOY-----LLKDNIIILKGYAL-----KVTYVNPCKTDMLEGVYGSKEKDIK---302
Db 1014 YNDNEYCOYNNSYCFDHSDLKNMFLNQHNSKLLTHSNKNSGFFNGINYESKHLANPEI 1073
Qy 303 -----IOKANP-----N 309
Db 1074 KTFAHNSYPIILNOGLINCNPLOCLGYDSNQRKNHVVYIKKNEYLKNGIINVLKREG 1133
Qy 310 LRALS-----ETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVTFPMND 356
Db 1134 LRKISTHNGKPFESFNMNDKQNYMEGLNIQDNVNNNNKESC-DNIKHMRTKSLNFSVRE 1192
Qy 357 K-GE--AIDKGNLVTDSKLVFGK-DDKEYTGEDKFNVEAIKED-GSMLFID---TKP 408
Db 1193 SYGEHKSLLDVYQECYVKNKLI--NKVNDKCY--EDNNNSYLNEDDNASMQFYETNSNP 1248
Qy 409 VNLSDM---KNYFNPSKSNKIYVRNPEFYLRGKISDKGGFNWE-----LRVNESV-----455
Db 1249 YIVDQENNMKNYV---NVLNNNSNYVDSKNYDKSKENAEKSDDLNENIHTLKD 1304
Qy 456 -----VDNLYIYGDHLIDNTRDFNITKLVNKGDIIDMW--GMKYKANGFPDKV-TMDMG 506
Db 1305 QKKKIQNNNEFISEQADINERISQEVVEYKEHEPL-WVINASNEBKSYEELIYSDMSS 1363
Qy 507 NYLQGYSDLNKAVGHYQFLYDNVKEVINIDPKGNTSIEYADGKSVVFNIDKRNNG 566
Db 1364 NRVTNKNYSMDNNVLLNEDNLLTTEKYKVQLEKENKMDMYETVEENINTIKTENTND 1423
Qy 567 FQGEIQEQHIYINGKEYTSFNDIKQIIDLTKLKIIVKVKDFARNT--TVK--EFILNKDTG 622
Db 1424 INEVR-----NEOKRSININD-----TWINHIIDEPNDTYNFIDKECVHNNENN 1472
Qy 623 EYSELKPHRVTVTIONKEMSSIVSEEDFLPVYKGELEKGYQDFGWEISGFGKKGAG 682
Db 1473 MYNSIEQVTFYHDTNRNHLVDK--NNQNF-----FEIEGLNELNFEKK--V 1516
Qy 683 YVINLSKD-----TFIKPVFKLIEEKEEENKPTF-----DVSKKKN 720
Db 1517 YIENNTKDDHKGDSKTSNLTSLRNTICKSENHNEKNENTYVVRKGEKGIKRKVKMKRN 1576
Qy 721 POVN-HSQLN-----ESHKEDLQREHSOKS 746
Db 1577 EKLNEENYNNIYKQNDNHRQNDITKKNDEEN 1609

RESULT 10
Q81H4 ID Q81H4 PRELIMINARY; PRT; 2849 AA.
AC Q81H4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF11_0392.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
```

RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angioli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrall B.,
RT "Genome sequence of the human malaria parasite Plasmodium
falciparum";
RL Nature 419:498-511 (2002).
DR EMBL; AE014841; AA035975.1; --
KW Hypothetical protein.
SQ SEQUENCE 2849 AA; 335916 MW; B5515D173D968813 CRC64;

Query Match 6.2%; Score 251.5; DB 5; Length 2849;
Best Local Similarity 19.1%; Pred. No. 0.046;
Matches 178; Conservative 162; Mismatches 361; Indels 231; Gaps 42;

QY . 8 SKFNLGNGKESLKKDITGVHHHQQEESIKEKSSFTIDRNISTIRDPEN--KDLKK- 64
DB 295 SYNNNNNNNNN--NNNSANVTNKSRRDKKYKVSFANQENHSYKFPQDKPSLKNK 353
QY 65 ---LIKK-KFREVDDFTSETGRMEYDYKYDDKGNIIAYD---DGTD-LEYETEKLDE 115
DB 354 KOTILKKIYEDELYKQYDOMEYEGEYSDS---LYDEMCESDDKIYVEDEKFN 409
QY 116 IKSIVGVLPSPKDGHEIFLG--KISNVSKNAKYVGNVYKSIETIKATYDFHSKTMTPD 173
DB 410 MVNK-----KVAKDINFWMNSKKVSKIDINL---KYQSDNRKMSNDSHVKS--SEN 459
QY 174 LVANINDIVDGLAFAGDMRLFYKNDQKAEIK-----TRMP-----EKIKETSEYP-- 221
DB 460 ILNNNKILNRVLDPKDISYVEEYQKRKEQAQNYHNVENPNYTDKNINDKNYPGD 519
QY 222 -----YVSSYGVNI-----ELGEGDLSKKNPNL-----TKMESGKI 253
DB 520 NINLNNNIRSNNTTVYANSTLSNEKNMNLNSNISDEVSSINKENLLNTNKNSENIT 579
QY 254 YSDSEKQOYLLKDNILKRGYALKVTTYNPGKT-----DMLBEG-----VYSKEDI 300
DB 580 IQNCDDNNKLVNDIVIDQEFILK--NYNVNKNKDYDYSTYDLSLKNNNFIKDIYTDNDV 637
QY 301 AKIQ-----KANPNLRALSETTIYADSRNVED-----CRSTQSVLMSALDGF 342
DB 638 INTESMGIYNETMKNKEISEPCIHNTNINYEYILTNMNTMNVGNENSDIILNKIEKT 697
QY 343 NIIR-----YQVFT-----FKMNDKGEAIDKGNLVTDSSKLVLFGKD--DKEYTGEDKEN 391
DB 698 NFENESKLYDVVDVINDIYKKNKEETINKIOQKCVDKVMYDFINNIDKETTINLDM-- 755
QY 392 VEAIKEDGSMFLFDITKPVNLSMD--KYVFNPSKSNKIYV----- 428
DB 756 -----DKXHLVNNKMNSTCDDKDTYLEMSKKEKINTFLKYLKMDVNSLSHLFQYFV 809
QY 429 -----RNPEFYLRG---KLSDKGGFNWELRVNESVVDNVLVYGLDLHIDNTRDNFK 476
DB 810 DREKDEEMKKQLFLIGDGEKKQMEFMYKINQNTQTLNKLGHESVOTNNEKKIQIE 869
QY 477 LNVKDGIDMDGMDKYKANGFPDKVTDMDGNVYLTQGYSDLNKAVGVHYQFLYDNVQPE 536
DB 870 -NIQTDIKDITKLYIKDMINKTSIDSVFF-----KSLSKDSYDLNKNKED 918
QY 537 VNIDPKGNTSIEYADGKVVFNINDKRNNGFDGEIOEHYINGKBYTSFNDIKQIIDKT 596
DB 919 IK---KNDIT-----YTKQEDMENKVD-----VTNVESNTETVNEVQKIND-- 956
QY 597 LNKIVVK-----DFARNITVKE-----FILNKDTGVESELXPHVTVTIQNGKMSST 645
DB 957 LKVKILLEIKGCVDNYSNNNNYNDDEQTAIILMDKNEYSKEKMYDYNLIDENRILSK 1016
QY 646 IVSEEDFILPVYKGELEKGYQFDGWEITSFGEGKDKAGYVNL-----S 688

DB 1017 LNDEENMKSNEYKNNKNSRW-----TVTFASLKSFEREMNLLKSHNERLRRIEKLYES 1072
QY 689 KDTFTKPVFKKLEEKKEENK---PTFDVSKKNDPQVNHSHOLNESHKEDLQRE--- 741
DB 1073 RDR-INKETIKMEKUKESQDRFLFIATERHIEKHNELNLSKKNED-MKYDLKKNNIKI 1130
QY 742 -HSQKSDSTKQVDTATVLDKNNISSKSTTNNPN 772
DB 1131 ALESQIDNNLNMISNNMEKNNNNNNNNNN 1162

RESULT 11
Q812V4
ID Q812V4 PRELIMINARY; PRT; 3381 AA.
AC Q812V4;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PFI0975C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255708; PubMed=12368867;
RA Hall N., Pain A., Beriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kethornou A.,
RA Knights A., Konfortov B., Kyes S., Lark N., Lawson D., Lennard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sultston J.E., Craig A., Newbold C., Barrell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531 (2002).
DR EMBL; AL929157; CAD51881.1; --
KW Hypothetical protein.
SQ SEQUENCE 3381 AA; 396148 MW; ODABOAFAB2EDB88 CRC64;

Query Match 6.2%; Score 251.5; DB 5; Length 3381;
Best Local Similarity 19.9%; Pred. No. 0.055;
Matches 193; Conservative 150; Mismatches 354; Indels 275; Gaps 45;

QY 18 EGSLLKDT-----TGVEHHQENESIKEKSSFTIDRNISTIRDPENKDLKK----- 64
DB 801 ENNMKNDFNNKNNTINDEQNDSEIYNSKQDDVSCDDSSV---NKDIRNNYEGIP 856
QY 65 LIKKKPREV-----DDFTSETGK-----RMEYDYKYDD---KGNIIAY 100
DB 857 VVEVQDEYNEQNDQDDDDDEDDGDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD 916
QY 101 DDGDTLEYETEKLDEIKSIYK-----VLT-----SPSK 128
DB 917 DDDTNNYEDEB-SFLSTNIFYGKNNSDNHNKNIININKYVDKDNIVLKSLSKSSSS 975
QY 129 DGHFEILGKISNVKNAKYVGNVYKSIETIK-----AT 161
DB 976 DG-----IKKSVSFSKYVYDLAKNNEPMMNFKSFDDITMSEMDDQDNINVTIPEHTFSQT 1031
QY 162 KYDFHSKTMTPDLYANINDIVDGLAFAGDMRLFYKNDQKAEIK-----NDOKKAEIKI- 207
DB 1032 NKKIYKKNNTSNTDLTINNYESTFEMKNLKEKKNQKDLNLSHKNQKEVIQIN 1091
QY 208 -----RMPEKIKETSEYPYVSSYGVNIEGEGDLSKKNPNL-----TKMESGKI 253

Db 4853 LLLKNE---NQIKTEKINKLNSDLNHFNTSLDKSAYQLSELNNEKKKKILLNELEEK 4909
Qy 665 YQFDGWEISGPEGKDGAGYVILNSKDTFIKPVFKKI--EEKKEENKPTFDVSKKD-NP 721
Db 4910 EKIISHLQKDFESNVE---VINELKE-FNEMLINKLKEEYKKEKDNVYD--KNSDHNY 4963
Qy 722 QVNHSQLNESHKREDLQREEHSQKSDSTKDVTATVLDKNISSKSTTNNPN 772
Db 4964 SNNNDNNNNNNSSNRSHSNNNNNNNSW-----NNMKSFHKCDIPN 5009

Search completed: January 28, 2004, 13:01:39
Job time : 48 secs